

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 19:45:08 ; Search time 779 Seconds

(without alignments)
10126.963 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 ggggaattgccttcaaat.....cctgaacctatcatgagc 1857

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: _geneseq1980s:*
2: _geneseq1990s:*
3: _geneseq2000s:*
4: _geneseq2001as:*
5: _geneseq2001bs:*
6: _geneseq2002s:*
7: _geneseq2003as:*
8: _geneseq2003bs:*
9: _geneseq2003cs:*
10: _geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1628	87.7	3082	4	AAH18299 Human CDN
2	636.4	34.3	751	4	AAH08535 Human CDN
3	549.2	29.6	1085	5	AAH29609 Human end
4	549	29.6	1118	4	AAH33270 DNA encod
5	549	29.6	1118	5	AAH29716 Human end
6	471.8	25.4	813	3	AAH21794 Human bre
7	283.6	15.3	2460	6	AAH19645 Human sec
8	142.6	7.7	344	3	AAH10622 Human sec
9	126.6	6.8	390	8	AAH179627 Human ova
10	124	6.7	543	8	AAH34058 Human end
11	117.8	6.3	452	4	AAH112455 Human end
12	117.8	6.3	452	4	AAH54163 Human foe
13	117.8	6.3	452	4	AAH31809 Probe #24
14	117.8	6.3	452	4	AAH43706 Human bre
15	117.8	6.3	452	4	AAH23908 Probe #23
16	117.8	6.3	452	4	AAH27874 Human bon
17	117.8	6.3	452	4	AAH24432 Human bra
18	117.8	6.3	452	4	AAH27457 Human liv
19	117.8	6.3	452	5	AAH102366 Probe #23
20	117.8	5.4	492	6	AAH502333 Human gen
21	101	5.4	492	4	AAH11820 Probe #17
22	101	5.4	492	4	AAH53513 Human foe
23	101	5.4	492	4	AAH31318 Probe #18

24	101	5.4	492	4	AAH43098
25	101	5.4	492	4	AAH23277
26	101	5.4	492	4	AAH27241
27	101	5.4	492	4	AAH01785
28	101	5.4	492	4	AAH26817
29	101	5.4	492	5	AAH01754
30	101	5.4	492	6	AAH01773
31	100.6	5.4	1204	6	AAH09522
32	100.6	5.4	2252	6	AAH61986
33	100.6	5.4	2255	9	AAH62727
34	95.6	5.1	809	4	AAH04673
35	95.6	5.1	2250	4	AAH15571
36	89.8	4.8	428	4	AAH13518
37	89.8	4.8	428	4	AAH55216
38	89.8	4.8	428	4	AAH34877
39	89.8	4.8	428	4	AAH44761
40	89.8	4.8	428	4	AAH24963
41	89.8	4.8	428	4	AAH28931
42	89.8	4.8	428	4	AAH03476
43	89.8	4.8	428	4	AAH28544
44	89.8	4.8	428	5	AAH03403
45	89.8	4.8	428	6	AAH503459

ALIGNMENTS

RESULT 1

AAH18299 standard; cDNA; 3082 BP.

AAH18299;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:18289.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-0018776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Claim 8; SEQ ID NO 18289; 2537bp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-

length cDNAs defined in the specification, where a primer set comprises:

(a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH1628 and AAH16633 to AAH18742 represent human cDNA sequences; AAB52446 to AAB58933 represent human amino acid sequences; and AAH11629 to AAH13332 represent oligonucleotides, all of which are used in the exemplification of the present invention

SQ Sequence 3082 BP; 628 A; 927 C; 916 G; 611 T; 0 U; 0 Other;

Query Match	Score	DB 4;	Length
87.7%;	1628;		3082;

Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	19	ATGATCCGAAATGTCGGAATCACTTCGACCGGGCTTGTCCTGTGTTCTTCCAAACCG	78
Db	83	ATGATCCGAAATGTCGGAATCACTTCGACCGGGCTTGTCCTGTGTTCTTCCAAACCG	142
QY	79	ACATTCGGGAAGTCAGCCTTAACTGTGCGGGGAACGACAGTGTCCATGTGCACACGGCCGAGGGA	138
Db	143	ACATTCGGGAAGTCAGCCTTAACTGTGCGGGGAACGACAGTGTCCATGTGCACACGGCCGAGGGA	202
QY	139	TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGTGTCTGCGCTCGCCACACCAACCA	198
Db	203	TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGTGTCTGCGCTCGCCACACCAACCA	262
QY	199	CGTGTGAGCTGAGCGCTCGGGCAGCTGCTCAAGTGAATTGACACACACGAGCTTGCTGTCC	258
Db	263	CGTGTGAGCTGAGCGCTCGGGCAGCTGCTCAAGTGAATTGACACACACGAGCTTGCTGTCC	322
QY	259	GCGGAGAGGCAACGCTGTGCTCCCTCTGTGTCAAGCTTCCATGATGACATCTAATTTTGG	318
Db	323	GCGGAGAGGCAACGCTGTGCTCCCTCTGTGTCAAGCTTCCATGATGACATCTAATTTTGG	382
QY	319	CACACCATTAAGTGTGGACAAGTGTGTAATATCAAGGCCAAGGTGAACCGGGCTTCAAC	378
Db	383	CACACCATTAAGTGTGGACAAGTGTGTAATATCAAGGCCAAGGTGAACCGGGCTTCAAC	442
QY	379	TCCAGCATGAGAGTGGGCAATCCAGGTGCTCGAGAGACTTGTCTCTGAGAAAGCAGTGG	438
Db	443	TCCAGCATGAGAGTGGGCAATCCAGGTGCTCGAGAGACTTGTCTCTGAGAAAGCAGTGG	502
QY	439	AATGTGTCCAAAGGCTTGGGCCACTTTCGTGGGCCGCCGAGAGATCAACCAAGTGAAGCTG	498
Db	503	AATGTGTCCAAAGGCTTGGGCCACTTTCGTGGGCCGCCGAGAGATCAACCAAGTGAAGCTG	562
QY	499	AAGCAGATCAACGCCCGCGGACAGAGAAGAGAAATGAGCAACAGTGTGGCGGCTGACCGC	558
Db	563	AAGCAGATCAACGCCCGCGGACAGAGAAGAGAAATGAGCAACAGTGTGGCGGCTGACCGC	622
QY	559	CGGCGCATATGCCCTTGTCTATGTCAAGACACCATTCAGAGACTTCTTGCCAACTGTGCCATT	618
Db	623	CGGCGCATATGCCCTTGTCTATGTCAAGACACCATTCAGAGACTTCTTGCCAACTGTGCCATT	682
QY	619	CAGGGCGATCTGGAGAGCAGAGACTGTAGCCGATGTGCCGCTAGAGAAAGCCGCTGTG	678
Db	683	CAGGGCGATCTGGAGAGCAGAGACTGTAGCCGATGTGCCGCTAGAGAAAGCCGCTGTG	742
QY	679	GAGAGTGTGAGAGTGTCTCTGCTCTCCCAACGCCCAATCAACACAGGGCAACACTTTTGGGGGC	738
Db	743	GAGAGTGTGAGAGTGTCTCTGCTCTCCCAACGCCCAATCAACAGGGCAACACTTTTGGGGGC	802
QY	739	CAGATCATGTGCTGTGATGAGAAATGTGGCACCATTTGACGCAAGCCGGGCTTCTGCGCTGCC	798

Db	803	CAGATCATGGCTGGATGAGAAATGTGACCAACATTGCAGCAGCCGGCTTCGGCCGTGC	862
QY	799	CACCTTCACGCTGMAAGGCCATTGAAATGTTCACTTCCGAGGCCCGTCCAGGTGGGAGC	858
Db	863	CACCTTCACGCTGMAAGGCCATTGAAATGTTCACTTCCGAGGCCCGTCCAGGTGGGAGC	922
QY	859	CGTGTGTGTCTCAAAAGCCATCGTGAAATAGTCCCTTCAACATATACATGAGAGTGGGCTG	918
Db	923	CGTGTGTGTCTCAAAAGCCATCGTGAAATAGTCCCTTCAACATATACATGAGAGTGGGCTG	982
QY	919	TGCGTGAAGGCTTATGCGCCAGAGAGGCTGAGACCCACCGGGCCCAATCAAGATGCTTTT	978
Db	983	TGCGTGAAGGCTTATGCGCCAGAGAGGCTGAGACCCACCGGGCCCAATCAAGATGCTTTT	1042
QY	979	ATGACCTTTTGGGCTCTGGAGCGAGATGACACAGGCCCAAGTGTGCTGACCTGATTCGAGCC	1038
Db	1043	ATGACCTTTTGGGCTCTGGAGCGAGATGACACAGGCCCAAGTGTGCTGACCTGATTCGAGCC	1102
QY	1039	CAGCCCGGCGCATGTGTGAGCGGCGGTATCCGAGAGGCCCATGTGCGAGAAAGAGATCCGCTG	1098
Db	1103	CAGCCCGGCGCATGTGTGAGCGGCGGTATCCGAGAGGCCCATGTGCGAGAAAGAGATCCGCTG	1162
QY	1099	GACAGGAAGTACATCGTGTCTCTGTAAACAGACAAAGAGGCCCTCGCTCGTCCCTGGGAC	1158
Db	1163	GACAGGAAGTACATCGTGTCTCTGTAAACAGACAAAGAGGCCCTCGCTCGTCCCTGGGAC	1222
QY	1159	CCTAGCAACCAAGGTGTATCCTGAGGCTACAATTAACGTCTCTCTTGAAGATGCTTGTGGCC	1218
Db	1223	CCTAGCAACCAAGGTGTATCCTGAGGCTACAATTAACGTCTCTCTCTTGAAGATGCTTGTGGCC	1282
QY	1219	AAGGACAACTGGGTGCTGTCTCTGGGAATACATGATAGTCCGCTGTACATCTGTGAAGAT	1278
Db	1283	AAGGACAACTGGGTGCTGTCTCTGGGAATACATGATAGTCCGCTGTACATCTGTGAAGAT	1342
QY	1279	GACAAAGTTCCTCTCTTCCACATGAGATGGTGTGATGTGAATGACAGCCCAAGCCCTTC	1338
Db	1343	GACAAAGTTCCTCTCTTCCACATGAGATGGTGTGATGTGAATGACAGCCCAAGCCCTTC	1402
QY	1339	CTGCTGCTCTTGGGACCTTGGCTCAGAGGCCAGAGTGGGACAAACATACCGGAGCGTGGAG	1398
Db	1403	CTGCTGCTCTTGGGACCTTGGCTCAGAGGCCAGAGTGGGACAAACATACCGGAGCGTGGAG	1462
QY	1399	CTTACTGACAGAGGTAGACGAGGACGAGGCCATTAACACGTCAACAGCCCTGCACCTCGGA	1458
Db	1463	CTTACTGACAGAGGTAGACGAGGACGAGGCCATTAACACGTCAACAGCCCTGCACCTCGGA	1522
QY	1459	GGTGCACCAAAAGCCCGACGACTTGATCTGTGATCTTGAGCAAT	1518
Db	1523	GGTGCACCAAAAGCCCGACGACTTGATCTGTGATCTTGAGCAAT	1582
QY	1519	GGGGAACCCCTATATGCAATCGGCTGAGGTGGGTCAAGCTGCGCCACACACCGAGAGCGCA	1578
Db	1583	GGGGAACCCCTATATGCAATCGGCTGAGGTGGGTCAAGCTGCGCCACACACCGAGAGCGCA	1642
QY	1579	GAGTACAGACGCGGAGAGACCCCTGTGCTCAGGCTTGTGCTCTTGCGCGAGAGGGGACACAG	1638
Db	1643	GAGTACAGACGCGGAGAGACCCCTGTGCTCAGGCTTGTGCTCTTGCGCGAGAGGGGACACAG	1702
QY	1639	CTGACCAAGTCTGCT	1654
Db	1703	CTGACCAAGTATCTT	1718

RESULT 2	
AAH08535	
ID	AAH08535 standard; cDNA; 751 BP.
XX	
AC	AAH08535;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA clone (5'-primer) SEQ ID NO:5370
XX	

PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236371P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239933P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.

PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

XX PA
XX FI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451936/48.
XX DR P-PSDB; AAU18380.
XX

PT Isolated polypeptide for treating, preventing and/or prognosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis.
XX

XX Claim 1; SEQ ID NO 109; 604bp; English.

XX Sequences AA629511-AA629736 represent cDNA molecules, which encode the
CC endocrine polypeptides of the invention. Endocrine polypeptides and their
CC associated polynucleotides of the invention are useful in the diagnosis,
CC treatment and prevention of various types of disorders in e.g. humans,
XX

CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
CC pathological condition can be determined by determining the presence or
CC absence of a mutation in an endocrine polypeptide. The treatable
CC disorders include autoimmune diseases such as rheumatoid arthritis,
CC hyperproliferative disorders such as neoplasms of the breast or liver,
CC cardiovascular disorders such as cardiac arrest, cerebrovascular
CC disorders such as cerebral ischemia, nervous system disorders such as
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at

Query Match 29.6%; Score 549.2; DB 5; Length 1085;
Best Local Similarity 98.6%; Pred. No. 1.9e-116;
Matches 554; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1093 CGCCTGACAGAGATCATGTGTCCTGTAAGACAGAGGTCCTCCGTCCTCC 1152
DB 247 CTCCTTCCAGAGATCATGTGTCCTGTAAGACAGAGGTCCTCCGTCCTCC 306
QY 1153 TGGAGACCTTACAGACAGAGTGTACCTGAGTACATTAACCTCTCTTGAAGATGCTT 1212
DB 307 TGGAGACCTTACAGACAGAGTGTACCTGAGTACATTAACCTCTCTTGAAGATGCTT 366
QY 1213 GTGGCCAGAGACAGATGAGGTGTCTCTCTGAGATCATGAGTCCGCTTGAACATCTG 1272
DB 367 GTGGCCAGAGACAGATGAGGTGTCTCTCTGAGATCATGAGTCCGCTTGAACATCTG 426
QY 1273 GAGAGTACAGATGCTCTCTCTTCCATGAGATGAGTGTGATGAGTGAAGCCAG 1332
DB 427 GAGAGTACAGATGCTCTCTCTTCCATGAGATGAGTGTGATGAGTGAAGCCAG 486
QY 1333 GCTTCTCTGCTCTCTCTGACCTGCTGACAGGCTGAGACAGACATACCGGAGC 1392
DB 487 GCTTCTCTGCTCTCTCTGACCTGCTGACAGGCTGAGACAGACATACCGGAGC 546
QY 1393 GTGGAGTGTGAGAGAGTGAAGAGAGACAGACATCAACAGTCAACGCTTGC 1452
DB 547 GTGGAGTGTGAGAGAGTGAAGAGAGACAGACATCAACAGTCAACGCTTGC 606
QY 1453 CTCGAGGTCAACAGAGCCCGAGACTTCGTGATCTCTGCTGAGGCGGAGCCTTGT 1512
DB 607 CTCGAGGTCAACAGAGCCCGAGACTTCGTGATCTCTGCTGAGGCGGAGCCTTGT 666
QY 1513 GACATGAGGAGCCCTATGTCATGCGCTGAGGTGCGTCAAGTCCACACAGAGAG 1572
DB 667 GACATGAGGAGCCCTATGTCATGCGCTGAGGTGCGTCAAGTCCACACAGAGAG 726
QY 1573 AGCCGAGATACAGAGCGGAGAGACCTCTGCTGACGCTTCTGCGCGAGAGAG 1632
DB 727 AGCCGAGATACAGAGCGGAGAGACCTCTGCTGACGCTTCTGCGCGAGAGAG 786
QY 1633 GACGAGTGAACAGTGTGCT 1654
DB 787 GACGAGTGAACAGTGTGCT 808

RESULT 4
AAS33270
ID AAS33270 strand: cDNA; 1118 BP.
XX AAS33270;
AC AAS33270;
XX
DT 04-DEC-2001 (first entry)
XX
DE DNA encoding human secreted protein, Seg ID No 229.

XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cyrostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing; ss.
XX
OS Homo sapiens.
XX
FN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US001347.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 25-SEP-2000; 2000US-0234984P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236377P.
 PR 29-SEP-2000; 2000US-0236386P.
 PR 29-SEP-2000; 2000US-0236399P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239355P.
 PR 13-OCT-2000; 2000US-0239357P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246522P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246538P.
 PR 08-NOV-2000; 2000US-0246603P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-451931/48.
 DR P-PSDB; AAU20561.
 XX
 PT New nucleic acids and polypeptides, useful for diagnosing, preventing or
 PT treating medical conditions.
 XX
 PS Claim 1; SEQ ID NO 229; 753pp; English.
 CC
 CC The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
 CC thrombosis), infections caused by bacteria, viruses and fungi and ocular
 CC disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
 CC and antibodies can also be used to promote wound healing, maintain organs
 CC before transplantation, and support cell culture of primary tissues.
 CC
 Query Match 29.6%; Score 549; DB 4; Length 1118;
 Best Local Similarity 96.9%; Pred. No. 2.1e-116;
 Matches 555; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 QY 1093 GCCTGACAGAGATATATGTGTCCTGTAGCAGAGAGTSCCTCTCGGTCCG 1152
 DB 246 CTCCTCCAGAGATATATGTGTCCTGTAGCAGAGAGTSCCTCTCGGTCCG 305
 QY 1153 TGGACCTTAGACACAGGTGATCTGAGCTACATTAAGTCTCTTGAAGATGCTT 1212
 DB 306 TGGACCTTAGACACAGGTGATCTGAGCTACATTAAGTCTCTTGAAGATGCTT 365
 QY 1213 GTGGCAGAGACACTGGGTGCTGCTCGAGATCACTGAGTCCGCTGTACTCTG 1272
 DB 366 GTGGCAGAGACACTGGGTGCTGCTCGAGATCACTGAGTCCGCTGTACTCTG 425
 QY 1273 GAGATGACAGTCTCTCTCTTCCACATGAGATGAGTGGCATGTGGATCAGCCCG 1332
 DB 426 GAGATGACAGTCTCTCTCTTCCACATGAGATGAGTGGCATGTGGATCAGCCCG 485
 QY 1333 GCCTTCCTGCTGCTCTCTCGACCTGAGGCGCAGATGGAGACAGACTACCGGACG 1392
 DB 486 GCCTTCCTGCTGCTCTCTCGACCTGAGGCGCAGATGGAGACAGACTACCGGACG 545
 QY 1393 GTGAGCTTAGTGACAGAGTAGACGAGCGCCATCTACCACTACACGCTGCTGCC 1452

DB 546 CTGGAGCTACTGACGAGGTAGCGAGAGAGAGCCATCTACCACTCACCACTCCCTGCC 605
QY 1453 CTGGAGGTACACAAAGCCCCAGAGACTTCGTGATCTCGGCTCGAGGGGAAAGCTTGT 1512
DB 606 CTCGGAGGTACACAAAGCCCCAGAGACTTCGTGATCTCGGCTCGAGGGGAAAGCTTGT 665
QY 1513 GACAAATGGGGAGCCCTATGTGATCGCGCTGAGAGGTGCGTACGCTGCCACACACCGAGAG 1572
DB 666 GACAAATGGGGAGCCCTATGTGATCGCGCTGAGAGGTGCGTACGCTGCCACACACCGAGAG 725
QY 1573 AGCCGAGTACGAGAGCGGAGAGACCTCTGCTGAGGCTTCTGCTGGCCGCGAGGGG 1632
DB 726 AGCCGAGTACGAGAGCGGAGAGACCTCTGCTGAGGCTTCTGCTGGCCGCGAGGGG 785
QY 1633 GACCACTGACCAAGTGTCTGCTGGTAAAGGTC 1665
DB 786 GACCACTGACCAAGTGTCTGCTGGTAAAGTCTC 818
RESULT 5
AAS29716
ID AAS29716 standard; cDNA; 1118 BP.
XX AAS29716;
AC
DT 21-NOV-2001 (first entry)
XX
DE Human endocrine polypeptide encoding cDNA SEQ ID NO 216.
XX
KW Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;
KW dog; chicken; sheep; immunosuppressive; antiarthritic; vasotrophic;
KW antirheumatic; antiproliferative; cytostatic; cardiatic; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; vitucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-fertility.
XX
OS Homo sapiens.
XX
FN WO200155364-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001308.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246529P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246603P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XI Rosen CA, Barash SC, Ruben SM.
 XX WPI: 2001-451936/48.
 DR P-PSDB; AAU18487.
 XX
 PT Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders of the endocrine system such as reproductive disorders,
 PT endocrine cancers and also for testing and detection e.g. diagnosis.
 XX
 PT
 PT
 XX
 PS Claim 1; SEQ ID NO 216; 604pp; English.
 XX
 PS Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the
 CC endocrine polypeptides of the invention. Endocrine polypeptides and their
 CC associated polynucleotides of the invention are useful in the diagnosis,
 CC treatment and prevention of various types of disorders in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
 CC pathological condition can be determined by determining the presence or
 CC absence of a mutation in an endocrine polynucleotide. The treatable
 CC disorders include autoimmune diseases such as rheumatoid arthritis,
 CC hyperproliferative disorders such as neoplasms of the breast or liver,
 CC cardiovascular disorders such as cardiac arrest, cerebrovascular
 CC disorders such as cerebral ischaemia, nervous system disorders such as
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,
 CC ocular disorders such as corneal infection, endocrine disorders such as
 CC premature labour and infertility, gastrointestinal disorders such as
 CC Crohn's disease, renal disorders such as glomerulonephritis and
 CC respiratory disorders such as asthma. The polypeptides can also be used
 CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, to regenerate tissues and in chemotaxis.
 CC The polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at

Query Match	29.6%	Score 549;	DB 5;	Length 118;
Best Local Similarity	96.9%;	Pred. No. 2.1e-116;		
Matches 555;	Conservative 4;	Mismatches 14;	Indels 0;	Gaps 0;

1093 CGCCTGACAGGAAGTACATCGTGTCTCTGTAAGCAGACAGAGGTGCCCCCTCTCCGTCCCC 1152

Db 246 CTTCTCCAGGAAGTACATCGTGTCTCTGTAAGCAGACAGAGGTGCCCTTTCGTCCTCC 305

[illegible]

1213 GTGGCCAAAGGACAACTGGGTGCTGTCTCGGAGATCACTCAAGTCCGCCCTGTAACTCTG 1272

Db 366 GTGCCAAGACAACCTGGTGCTGTCTCCGAGATCAGTCAGTCCGCTGTACACTCTG 425

1273 GAGATGACAAGTTCTCTCTCTCCACATGAGATGGTGTGATGTGATGCACCCAG 1332

Db 426 GAGGATGACAA GTTCCCTCTCTCCACATGGAGATGGTGGTCATGTGGATGCAGACCAG 485

[illegible]

1393 GTCGACCTTACCTGCACAGGAGACGAGAGACGACGCAATCTTACCAACGTTACGAGGCTGAC 1455

Db 546 GTGGAGCTAGTGCAGCAGGTAGACGAGGACGACGCCATCTACCACGTCACCAACCCCTGCC 605

1453 CTCGAGGTCACACAAAGCCCCAGGACTTCGTGATCTGGCCTCGAGGCGGAAGCCTTGT 1512

Db 606 CTCGAGGTCACACAAAGCCCCAGACTTCGTGATCCTGGCCCTCGAGGGCGAAGCCTTGT 665

1513 GACAAATGGGAGACCCCTAIGTCATCGCGCTGAGGTCGGTCAAGCTGCCCACACACCGAGAG 13/4

1 5 3 2

Db 726 ACGCCAGAGTACAGACGCGGAGAGACCTCTGCTCTCAGGCTTCTGCTCTGCGCGGAGG 785

1633 GACCACTGACCAAGTGTCTGGGTAGGTC 1665

Db 786 GACCAGTGACCAAGGTAGCTGTAGTACTC 818

RESULT 6

AAF21794 standard; DNA; 813 BP.

AC AAF21794;
XX

DT 27-MAR-2001 (first entry)
XX

[illegible]

KW neurotropic; neurprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antiulcer; vulnerrary; anticonvulsant;

KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW antibacterial; antitubercal; antiparasitic; catarrhal; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;

KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disease; wound healing; neurological disease. ds

XX
05 Homo sapiens.

AA
PN WO200055173-A1.

PD 21-SEP-2000.

CC the proteome of a tissue or a cell type. The polynucleotide is useful for
CC creating knockin humanised animals (pigs) or transgenic animals (mice or
CC rats) to model human disease, and for somatic or germ-line gene therapy,
CC and further for generating hybridisation probes useful in mapping the
CC naturally occurring genomic sequence. This polynucleotide sequence
CC represents the DNA of a human secreted protein of the invention
XX

Sequence 2460 BP; 501 A; 717 C; 753 G; 489 T; 0 U; 0 Other;

Query Match 15.3%; Score 283.6; DB 6; Length 2460;

Best Local Similarity 98.6%; Pred. No. 3.1e-55;

Matches 286; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1365 GCCAGAGGAGGAGCAACGACTACCGGAGGAGTGAAGTGAAGAGAGAGCA 1424

DB 2460 GCCAGAGGAGGAGCAACGACTACCGGAGGAGTGAAGTGAAGAGAGAGCA 2401

QY 1425 CGCCATCTACGACGTCACACGCTGCTGCTGAGAGTCAACAAAGCCCAAGACTTGT 1484

DB 2400 CGCCATCTACGACGTCACACGCTGCTGCTGAGAGTCAACAAAGCCCAAGACTTGT 2341

QY 1485 GATCTGAGCTTCGAGGCGGAGGAGCTTGTGACATGGGAGACCTTATGTCATGCGCTGAG 1544

DB 2340 GATCTGAGCTTCGAGGCGGAGGAGCTTGTGACATGGGAGACCTTATGTCATGCGCTGAG 2281

QY 1545 GTCCGTACGCTGACGACACGAGAGAGCGCAGAGTACAGACGCGAGAGACCTCTG 1604

DB 2280 GTCCGTACGCTGACGACACGAGAGAGCGCAGAGTACAGACGCGAGAGACCTCTG 2221

QY 1605 CTCAGGCTTTCGCTCTGCGCGAGGAGGAGGACCTGACCAAGTGTCTGCT 1654

DB 2220 CTCAGGCTTTCGCTCTGCGCGAGGAGGAGGACCTGACCAAGTGTCTGCT 2171

RESULT 8

AACT10622

ID AACT10622 standard; cDNA; 344 BP.

AC AACT10622;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 14697.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

PA (GIST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 14697; 71pp + Sequence Listing; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors

Sequence 344 BP; 65 A; 91 C; 127 G; 59 T; 0 U; 2 Other;

Query Match 7.7%; Score 142.6; DB 3; Length 344;

Best Local Similarity 98.2%; Pred. No. 5.1e-23;

Matches 164; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

QY 19 ATGATCCAGAAATGTCGAAATCACCCTGCGACGCGGCTTGGCTGTGTTCTCAACCGC 78

DB 176 ATGATCCAGAAATGTCGAAATCACCCTGCGACGCGGCTTGGCTGTGTTCTCAACCGC 235

QY 79 ACATCCCGGAAGTCAGCTTACGTCGCGGGAACGACATGCCATGCGACGCGGAGGA 138

DB 236 ACATCCCGGAAGTCAGCTTACGTCGCGGGAACGACATGCCATGCGACGCGGAGGA 295

QY 139 TACCGG-AACCCGACGAGGTGACAGATGAG-CCAGCTGTGTGCTGCC 183

DB 236 TACCGGAACCCGACGAGGTGACAGATGAG-CCAGCTGTGTGCTGCC 342

RESULT 9

ABL79627/c

ID ABL79627 standard; cDNA; 390 BP.

AC ABL79627;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:2605.

DE Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US017756.

XX 26-MAY-2000; 2000US-0207484P.

PA (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

PT Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.

PS Claim 1; SEQ ID NO 2605; 489pp; English.

CC The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably

CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
SQ Sequence 390 BP; 96 A; 100 C; 86 G; 108 T; 0 U; 0 Other;
Query Match 6.8%; Score 126.6; DB 6; Length 390;
Best Local Similarity 87.9%; Pred. No. 2.5e-19;
Matches 138; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1644 CAACTGCTGCTGGGTTAGGCTCTCCCTGACTGAGCTGGTCTGGCAAGTGGCTTATTC 1703
Db 357 CTAGCTTCATTTGGGTTAGAGTCTCCCGAGCGAGCTGCTCGGCAAGTGGCGCTCATTC 238
QY 1704 CTGAGGAGCTGCAATCCAGTCCAGAGGGTCCAGAGCGAGGTTGATGAAACTAGC 1763
Db 297 GTGAGAGCTGCAATCCAGTCCAGAGGGTCTGGAGCGAGGTTGATGAAACTAGC 238
QY 1764 TGGAGAGCACCTGAGTACTCTTAAAGCAATCCCGTG 1800
Db 237 TGGAGAGCACCCAGTACTCTTAAAGCAATCCCGTG 201
RESULT 10
ACH34058
ID ACH34058 standard; cDNA; 543 BP.
XX ACH34058;
AC 13-OCT-2003 (first entry)
XX 13-OCT-2003 (first entry)
DE Human endothelial cell cDNA #2191.
XX
XX Human: ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
OS Homo sapiens.
XX
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridisation probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX Claim 1; SEQ ID NO 21270; 44bp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of

CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspro.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 543 BP; 141 A; 136 C; 155 G; 108 T; 0 U; 3 Other;
Query Match 6.7%; Score 124; DB 8; Length 543;
Best Local Similarity 75.0%; Pred. No. 1.1e-18;
Matches 195; Conservative 0; Mismatches 60; Indels 5; Gaps 3;
QY 1601 TCTGCTCAGGCTTCTGCTCTGCGCGAGGGGACACAGCTGACCAAGTCTGCTGGGTTA 1660
Db 34 TCGCACAGAGAGCTGGCGCGCTGACAGTGTGCTGAGTGTCTCCAGTGCACATGGGTTA 93
QY 1661 GGGTCTCCCTGACTGAGCTGGTCTGCGCAAGTGGCTTCTTCTGGGGGCTGCAATCCA 1720
Db 94 GGGTCTCCACACACGAGAGCTGGTCTCAAGAGTGTGCTCAAT-GGGGCTGCAACTG 152
QY 1721 GGTCAAGGGTCCAGAGCGAGCGAGGCTG--AATGAAACCTAGCTGAGGACACCTGA 1777
Db 153 GGTGTAAGGGTCCAGAGCGAGCGAGGAGGAAAGTGAAGTGAAGCTGAGGAGACCTGA 212
QY 1778 GTACTCTTAAGCAATCCCGTGGCCAAATCAACAGCGATTGGATTACCTTCAAGACA 1837
Db 213 GTACTCTT-AAAGCAATCCCGTGGACCAATCAATGCAATTTGGATTATCATCGAGCA 271
QY 1838 CCTGAACCTTATCATGAGC 1857
Db 272 CCTGAACCTTATCATGAGC 291
RESULT 11
AA112455
ID AA112455 standard; DNA; 452 BP.
XX AA112455;
AC 12-OCT-2001 (first entry)
XX
XX 12-OCT-2001 (first entry)
DE Probe #2388 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX MO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-0063236P.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
XX Claim 25; SEQ ID NO 2388; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;
XX
Query Match 6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;
XX
QY 1644 CAAATGCTGCTGGGTTAGGGTCTCCCTGACAGACTGCTCGGCAAGTGGCTTCTATTC 1703
DB 70 CTAATGTCACATGGGTTAGGGTCTCCCTGACAGACTGCTCGGCAAGTGGTGTCCA-TC 128
QY 1704 CTGGGGGCTCGAATCCAGTCCAAAGGTCGACGAGCGACGGTTGGAATGAAAACCTAGC 1763
DB 129 GTGGGGGCTCGAATCCAGTCCAAAGTCTTCTGAGCGATGTTGGAATGAAAACCTAGC 188
QY 1764 TGGAGGACACCTGAGTACTCTTAAGCAATCCCGTGGCCAAATCAACAGCCGATTTGG 1823
DB 189 TGGAGGACACCGGAGTACTCTTAAGCAATCTCGTGTGAGTAAGAGGGGAGCTAGGA 248
QY 1824 A 1824
DB 249 A 249
XX
RESULT 12
ABA54163
ID ABA54163 standard; DNA; 452 BP.
XX
AC ABA54163;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #2468.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0068408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
XX Claim 1; SEQ ID NO 2468; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;
XX
Query Match 6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;
XX
QY 1644 CAAATGCTGCTGGGTTAGGGTCTCCCTGACAGACTGCTCGGCAAGTGGCTTCTATTC 1703
DB 70 CTAATGTCACATGGGTTAGGGTCTCCCTGACAGACTGCTCGGCAAGTGGTGTCCA-TC 128
QY 1704 CTGGGGGCTCGAATCCAGTCCAAAGGTCGACGAGCGACGGTTGGAATGAAAACCTAGC 1763
DB 129 GTGGGGGCTCGAATCCAGTCCAAAGTCTTCTGAGCGATGTTGGAATGAAAACCTAGC 188
QY 1764 TGGAGGACACCTGAGTACTCTTAAGCAATCCCGTGGCCAAATCAACAGCCGATTTGG 1823
DB 189 TGGAGGACACCGGAGTACTCTTAAGCAATCTCGTGTGAGTAAGAGGGGAGCTAGGA 248
QY 1824 A 1824
DB 249 A 249
XX
RESULT 13
AAI3809
ID AAI3809 standard; DNA; 452 BP.
XX
AC AAI3809;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #2495 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0068408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.


```

XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI  WPI; 2001-468897/53.
XX
XX
PT  Human genome-derived single exon nucleic acid probes useful for analyzing
PT  gene expression in human placenta.
XX
XX  Claim 25; SEQ ID NO 2495; 654bp; English.
XX
CC  The present invention relates to single exon nucleic acid probes (SENP).
CC  The present sequence is one such probe. The probes are useful for
CC  producing a microarray for predicting, measuring and displaying gene
CC  expression in samples derived from human placenta. The probes are useful
CC  for antenatal diagnosis of human genetic disorders
XX
SQ  Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match      6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY  1644 CAAGTGTGCTGGGTTAGGCTCTCCCTGACTGAGCTGTCTCCGGCAAGTGGCTTCTATTC 1703
    |||||
DB  70  CTAGTGCACCTGGGTTAGGCTCTCCCTGACCGAGCTGTCTCCGCAAGTTGTGCCA-TC 128

QY  1704 CTGGGGGCTCGAATCCAGTCAAAAGGCTCCGAGAGCGATGGAATGAAAACCTAGC 1763
    |||||
DB  129  GTGGGGGCTCGAATCCAGATCCAGATCGAAGTCTTTCTGAGCGATGTGGAAATGAAAACCTAGC 188

QY  1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
    |||||
DB  189  TGGAGGACACCCGAGTACTCTTAAAGCAATCTGTGTGATGAAGAGGAGCTAGGA 248

QY  1824 A 1824
DB  249 A 249

RESULT 14
ABA43706
ID  ABA43706 standard; DNA; 452 BP.
XX
XX  ABA43706;
XX
XX  01-FEB-2002 (first entry)
XX
DE  Human breast cell single exon nucleic acid probe #2401.
XX
XX  Human; microarray; single exon probe; gene expression; breast; disease;
XX  cancer; ss.
XX
OS  Homo sapiens.
XX
XX  WO200157271-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US000662.
XX
XX  04-FEB-2000; 2000US-0180312P.
XX  26-MAY-2000; 2000US-0207456P.
XX  30-JUN-2000; 2000US-00608408.
XX  03-AUG-2000; 2000US-00632366.
XX  21-SEP-2000; 2000US-0234687P.
XX  27-SEP-2000; 2000US-0236359P.
XX  04-OCT-2000; 2000GB-00024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
XX  WPI; 2001-496933/54.

```

```

XX  New spatially-addressable set of single exon nucleic acid probes, useful
PT  for measuring gene expression in sample derived from human breast,
PT  comprises number of single exon nucleic acid probes.
XX
XX
PS  Claim 1; SEQ ID NO 2401; 327bp + Sequence Listing; English.
XX
XX  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human breast and BT 474 cells. The method involves contacting the
CC  probes with a collection of detectably labelled nucleic acids derived
CC  from mRNA of human breast, and then measuring the label bound to each
CC  probe of the microarray. The probes are useful for verifying the
CC  expression of regions of genomic DNA predicted to encode proteins. They
CC  are useful for gene discovery, and for determining predisposition and/or
CC  prognosing breast disease. Gene expression analysis is useful for
CC  assessing the toxicity of chemical agents on cells. The microarray of
CC  this invention presents a far greater diversity of probes for measuring
CC  gene expression, with far less bias than expressed sequence tag
CC  microarrays. The method is suitable for rapid production of functional
CC  information from genomic sequence. The present sequence is a single exon
CC  nucleic acid probe of the invention. Note: The sequence data for this
CC  patent did not form part of the printed specification, but was obtained
CC  in electronic format directly from WIPO at
    ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

Query Match      6.3%; Score 117.8; DB 4; Length 452;
Best Local Similarity 81.8%; Pred. No. 2.8e-17;
Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY  1644 CAAGTGTGCTGGGTTAGGCTCTCCCTGACTGAGCTGTCTCCGCAAGTGGCTTCTATTC 1703
    |||||
DB  70  CTAGTGCACCTGGGTTAGGCTCTCCCTGACCGAGCTGTCTCCGCAAGTTGTGCCA-TC 128

QY  1704 CTGGGGGCTCGAATCCAGTCAAAAGGCTCCGAGAGCGATGGAATGAAAACCTAGC 1763
    |||||
DB  129  GTGGGGGCTCGAATCCAGATCCAGATCGAAGTCTTTCTGAGCGATGTGGAAATGAAAACCTAGC 188

QY  1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
    |||||
DB  189  TGGAGGACACCCGAGTACTCTTAAAGCAATCTGTGTGATGAAGAGGAGCTAGGA 248

QY  1824 A 1824
DB  249 A 249

RESULT 15
ABA23908
ID  ABA23908 standard; DNA; 452 BP.
XX
XX  ABA23908;
XX
XX  23-JAN-2002 (first entry)
XX
DE  Probe #2374 for gene expression analysis in human heart cell sample.
XX
XX  Human; gene expression; heart; microarray; vascular system; probe;
XX  cardiovascular disease; hypertension; cardiac arrhythmia;
XX  congenital heart disease; ss.
XX
OS  Homo sapiens.
XX
XX  WO200157274-A2.
XX
XX  09-AUG-2001.
XX
XX  30-JAN-2001; 2001WO-US000666.
XX
XX  04-FEB-2000; 2000US-0180312P.
XX  26-MAY-2000; 2000US-0207456P.

```

PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.

XX Claim 1; SEQ ID NO 2374; 530bp; English.

CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 452 BP; 115 A; 88 C; 123 G; 126 T; 0 U; 0 Other;

XX Query Match 6.3%; Score 117.8; DB 4; Length 452;

XX Best Local Similarity 81.8%; Pred. No. 2.8e-17;

XX Matches 148; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 1644 CAAGTGTGCTGGGTTAGGGTCTCCCTGACTGAGCTGGTCTGGCAAGTGGCTTCTATTTC 1703
Db 70 CTAGTGCACCTGGGTTAGGGTCTCCCTGACGAGCTGGTCTGGCAAGTGTGTCCA-TC 128

QY 1704 CTGGGGGGCTCGAATCGAGTCAAGGGTCGACGAGCGACGGTTGGAAATGAAAACTAGC 1763
Db 129 GTGGGGGGCTCGAATCGAATCGAAGTGTCTTGGAGCGATGGTGGAAATGAAAACTAGC 188

QY 1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGTGGCCAAATCAACAGCCGATTGG 1823
Db 189 TGGAGGACACCGAGTACTCTTAAAGCAATCTCTGTGAGTAAAGGAGGAGCTAGGA 248

QY 1824 A 1824
Db 249 A 249

Search completed: August 20, 2004, 21:31:04
Job time : 787 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 19:45:09 ; Search time 7436 Seconds

(without alignments)
10824.101 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 gtgaattgccttcaaat.....cctgaacctatcatgagc 1857

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rnd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1857	100.0	1857	AF416921	AF416921 Homo sapi
2	1839	99.0	6359	AB014607	AB014607 Homo sapi
3	1633.6	88.0	1818	AF416922	AF416922 Homo sapi
4	1628	87.7	3082	AX883384	AX883384 Sequence
5	1628	87.7	3082	BD160291	BD160291 Primer fo
6	1628	87.7	3082	AK023937	AK023937 Homo sapi
7	1450	78.1	2875	AK091112	AK091112 Homo sapi
8	1230	66.2	2678	AF416923	AF416923 Mus muscu
9	1230	66.2	5606	BC042492	BC042492 Mus muscu
10	1182.2	63.7	5415	AK122350	AK122350 Mus muscu
11	730.8	39.4	2455	BC001517	BC001517 Homo sapi
12	636.4	34.3	751	AX870465	AX870465 Sequence
13	636.4	34.3	751	BD150527	BD150527 Primer fo
14	434.8	23.4	1689	AB078618	AB078618 Mus muscu
15	433.6	23.3	1725	AB040609	AB040609 Rattus no
16	425	22.9	3019	BC025852	BC025852 Mus muscu
17	405	21.8	1820	AB078619	AB078619 Homo sapi
18	405	21.8	2077	AK122960	AK122960 Homo sapi
19	211.4	11.4	159681	AC036239	AC036239 Homo sapi
20	209.8	11.3	122146	AC011736	AC011736 Homo sapi
21	209.8	11.3	132033	AC099796	AC099796 Homo sapi
22	158.4	8.5	62485	AL590093	AL590093 Human DNA
23	156.8	8.4	159681	AC036239	AC036239 Homo sapi
24	143.2	7.7	2242	AX834086	AX834086 Sequence
25	143.2	7.7	2242	AK096315	AK096315 Homo sapi
26	143.2	7.7	246011	AL929585	AL929585 Mouse DNA
27	142.6	7.7	344	AX898834	AX898834 Sequence
28	142.6	7.7	344	BD034367	BD034367 Rattus no
29	140.8	7.6	252616	AC122593	AC122593 Rattus no
30	134.6	7.2	169770	AC021763	AC021763 Homo sapi
31	134.6	7.2	174741	AC040977	AC040977 Homo sapi
32	132.6	7.1	162914	AC008813	AC008813 Homo sapi
33	131.4	7.1	77001	AL513013	AL513013 Human DNA
34	131.4	7.1	154361	AC092061	AC092061 Homo sapi
35	131.4	7.1	167331	AC073115	AC073115 Homo sapi
36	131.4	7.1	177492	AC073350	AC073350 Homo sapi
37	130.6	7.0	147181	AL157931	AL157931 Human DNA
38	130.4	7.0	172274	AP000924	AP000924 Homo sapi
39	130.4	7.0	174084	AC084395	AC084395 Homo sapi
40	130.4	7.0	174311	AP002788	AP002788 Homo sapi
41	130.4	7.0	190592	AC015806	AC015806 Homo sapi
42	129.8	6.9	170470	AC092364	AC092364 Homo sapi
43	128.2	6.9	180510	AC022415	AC022415 Homo sapi
44	125.8	6.8	92941	AC011313	AC011313 Homo sapi
45	125	6.7	189278	AC055787	AC055787 Homo sapi

ALIGNMENTS

RESULT 1
AF416921
LOCUS AF416921 1857 bp mRNA linear PRI 13-DEC-2001
DEFINITION Homo sapiens brown fat inducible thioesterase 1 mRNA, complete cds,
alternatively spliced.
ACCESSION AF416921
VERSION AF416921.1 GI:17646236
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1857) Schilbach,S.L., Yu,X.X., Goddard,A.D.,
Adams,S.H., Chui,C., Dowd,P., Colman,S. and Lewin,D.A.
Grimaldi,C., Lee,J.,

TITLE

BFIT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adipose Tissue. Cloning, Organization of the Human Gene, & Assessment of a Potential Link to Obesity

JOURNAL

Biochem. J. (2002) In press

AUTHORS

Adams, S.H., Chui, C., Schilbach, S.L., Yu, X.X., Goddard, A.D., Grimaldi, C., Lee, J., Dowd, P., Colman, S. and Lewin, D.A.

TITLE

Direct Submission

JOURNAL

Submitted (07-SEP-2001) Endocrinology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA

FEATURES

source

Location/Qualifiers

1. 1857

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="1"

19. 1842

/note="BFIT1; acyl-CoA thioesterase; acyl-CoA hydrolase; alternatively spliced"

/codon_start=1

/product="brown fat inducible thioesterase 1"

/protein_id="AA140937.1"

/db_xref="GI:17646237"

/translation="MIQVGNHLRRLGLASVFSNRTSRKSLRAGNDSAMADGEYRNP

TEVMSQVLPCHTNORGLSYGLLKWIDTACASRRHAGCCPVTSMDIDYEHT

ISVGQVNIKAKVNAENSMVGIQVASEDISEKQNNVKKALATFVARREITVEKI

KOIPRTREKKEHSVAARRRMRLVADDTIKDLANALIGDLESRCSRMVPAEKT

RVESEVLVPHANQNTFGQILAMMENAVTILASLCAHPTLKAIEMHFGPS

QVGDVLKALVNNAFKSMEEVCEVAEQEAETRRHINSAPMTFVLLADDDPOL

LPWIRPQDEDRRYREASARKIRLDKRYVSCQTPVLSVPMDPNQVYLSTNNY

SLKMLVAKDNWVLSSEISQVRLYLEDKFLSPHMMVAVHDAQAFLLSDLPOR

EMDKHRSVELVQVDEDDAIYHVTSPALGGHTKPDVILASRRKPCDNDPVIAT

RSVILPHTREPEYRGRGLCSRCLEMEGQOLTKCMVRYSLTILVASGFYSWGLE

SRSKRRBDGMNKLGGHLSLTKAIPVAKINSRGTQDT"

ORIGIN

Query Match 100.0%; Score 1857; DB 9; Length 1857;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1857; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGGAATGCGCTTCAATGATCGAATGTGGAAATACCTGCGAGCGGGCTTGGCC 60
DB 1 GTGGAATGCGCTTCAATGATCGAATGTGGAAATACCTGCGAGGGGCTTGGCC 60
QY 61 TCTGTGTTCTCCAAACCGGACATCCCGGAAGTCAAGCTTACGTCGTCG 120
DB 61 TCTGTGTTCTCCAAACCGGACATCCCGGAAGTCAAGCTTACGTCGTCGTCG 120
QY 121 ATGCGACGCGGAGGAGTACCGGAACCCACGAGAGTGCAGATGAGCGAGTGTGCTG 180
DB 121 ATGCGACGCGGAGGAGTACCGGAACCCACGAGAGTGCAGATGAGCGAGTGTGCTG 180
QY 181 CCTGCGCACACCAACCAACGCTGTGAGCTGCGGCAAGCTGCTCAAGTGATTAAC 240
DB 181 CCTGCGCACACCAACCAACGCTGTGAGCTGCGGCAAGCTGCTCAAGTGATTAAC 240
QY 241 ACCAGGCTTGTGCTGCGGAGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 241 ACCAGGCTTGTGCTGCGGAGAGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 GATGACATCTATTGAGCAGACCATTAAGTGTGAGCAAGTGTGATATCAAGGCCAAG 360
DB 301 GATGACATCTATTGAGCAGACCATTAAGTGTGAGCAAGTGTGATATCAAGGCCAAG 360
QY 361 GTGAAACCGGCGCTTCACTCCAGCATGAGGTGTGAGTGTGAGTGTGAGTGTGAG 420
DB 361 GTGAAACCGGCGCTTCACTCCAGCATGAGGTGTGAGTGTGAGTGTGAGTGTGAG 420
QY 421 TGGCTGTGAGAGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 480
DB 421 TGGCTGTGAGAGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 480
QY 481 ATCAACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540
DB 481 ATCAACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540

```

```

DB 481 ATCAACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540
QY 541 AGTGTGCGGCTGAGCGCGCGGCGCATGCGCTTGTCTATGACAGACCATCAAGACCTC 600
DB 541 AGTGTGCGGCTGAGCGCGCGGCGCATGCGCTTGTCTATGACAGACCATCAAGACCTC 600
QY 601 CTGGCCAACTGGCCATTCAAGGCGCATCTGAGAGCAGAGATCTGAGCCGATGTGCGG 660
DB 601 CTGGCCAACTGGCCATTCAAGGCGCATCTGAGAGCAGAGATCTGAGCCGATGTGCGG 660
QY 661 GCTGAGAAAGACCGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
DB 661 GCTGAGAAAGACCGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
QY 721 GGCAACACCTTTGGGGGCGCAATCATAGCGCTGAGTGTGAGTGTGAGTGTGAGTGTG 780
DB 721 GGCAACACCTTTGGGGGCGCAATCATAGCGCTGAGTGTGAGTGTGAGTGTGAGTGTG 780
QY 781 AGCGGCGCTGCGCGTGGCCCAACCTTACGCTGAGAGGCGCATTTGAATGTTCATTCGAGGC 840
DB 781 AGCGGCGCTGCGCGTGGCCCAACCTTACGCTGAGAGGCGCATTTGAATGTTCATTCGAGGC 840
QY 841 CCGTCCAGAGTGTGCGGCGCGCATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 900
DB 841 CCGTCCAGAGTGTGCGGCGCGCATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 900
QY 901 AGCATGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 960
DB 901 AGCATGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 960
QY 961 CACATCAACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1020
DB 961 CACATCAACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1020
QY 1021 CTGCGCTGATTTGGGCGCGCGCATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1080
DB 1021 CTGCGCTGATTTGGGCGCGCGCATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1080
QY 1081 AGAAAGAGATTCGCTGTGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1140
DB 1081 AGAAAGAGATTCGCTGTGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1140
QY 1141 CTCTCGTCCCTGTGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1200
DB 1141 CTCTCGTCCCTGTGAGCAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1200
QY 1201 TTGAAGATGTGTTGGCCAGAGCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1260
DB 1201 TTGAAGATGTGTTGGCCAGAGCACTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1260
QY 1261 CTGTACATCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1320
DB 1261 CTGTACATCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1320
QY 1321 GATGAGAGCGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 1321 GATGAGAGCGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1381 CACTACCGGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1440
DB 1381 CACTACCGGAGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1440
QY 1441 ACCAGCGCTGCTGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1500
DB 1441 ACCAGCGCTGCTGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1500
QY 1501 CGGAAGCTTGTGAGCAATGGGAGACCCCTATGCTATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1560
DB 1501 CGGAAGCTTGTGAGCAATGGGAGACCCCTATGCTATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1560
QY 1561 ACAACCGAGAGAGCGCCAGATGACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 ACAACCGAGAGAGCGCCAGATGACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620

```

Db 1561 ACAACCCGAGAGACGCCAGAGTACAGACGCCGAGAGACCTCTGCTCAGGCTTCTGCTC 1620

QY 1621 TGCGCGAGGGGAGACCACTGACCAAGTCTGCTGAGTTAGGCTCTCCCTGACTGAGCTG 1680

Db 1621 TGCGCGAGGGGAGACCACTGACCAAGTCTGCTGAGTTAGGCTCTCCCTGACTGAGCTG 1680

QY 1681 GTCTCGGCAAGTGGCTTCTATTCCTGAGGGGCTGMAATCCAGGTCMAAGGGTCCGAGAGC 1740

Db 1681 GTCTCGGCAAGTGGCTTCTATTCCTGAGGGGCTGMAATCCAGGTCMAAGGGTCCGAGAGC 1740

QY 1741 GACGGTTGGAATGAAAGTCTGAGAGACACCTGAGTCTCTTAAAGCAATCCCGCTG 1800

Db 1741 GACGGTTGGAATGAAAGTCTGAGAGACACCTGAGTCTCTTAAAGCAATCCCGCTG 1800

QY 1801 GCCAAATCAACAGCCGATTGGATACCTTCAAGACACCTGAAACCTTATCATGAGC 1857

Db 1801 GCCAAATCAACAGCCGATTGGATACCTTCAAGACACCTGAAACCTTATCATGAGC 1857

RESULT 2

AB014607 6359 bp mRNA linear PRI 06-FEB-1999

LOCUS Homo sapiens mRNA for KIAA0707 protein, partial cds.

DEFINITION AB014607

ACCESSION AB014607.1 GI:33272227

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (sites)

1 Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.

TITLE Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro

JOURNAL DNA Res. 5 (3), 169-176 (1998)

MEDLINE 9734811

REFERENCE

AUTHORS

2 (bases 1 to 6359)

Ohara, O., Suyama, M., Nagase, T. and Ishikawa, K.

TITLE Direct Submission

JOURNAL Submitted (26-May-1998) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

FEATURES

source

1..6359

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="HH00286"

/sex="male"

/tissue_type="brain"

/clone_id="pBluescriptII SK plus"

/dev_stage="adult"

1..6359

/gene="KIAA0707"

<1..1895

/gene="KIAA0707"

/note="HH00286 cDNA clone for KIAA0707 has a 1-bp deletion at 1560 of the sequence of KIAA0707."

/codon_start=3

/product="KIAA0707 protein"

/protein_id="BAA31682.1"

/db_xref="GI:33272228"

/translation="PGCCSGDQVLSIGRCFPPPPAMTQNVGNHRLRGASVFSNETSRKSLRAGNDSAMADGEGYRNPTEVQMSQVLPCHTNQGELSVQQLKWDITTAQLSARRAGCCPTVATMDIYFEHTISVGOVNAIKXNRAPNSMVEGIVASPDICSEKQNNVCKALATFVAREITVKLQKITPREBEKMEHSVAERRMILVADTIKLLANCAIGDLESRDSCRMVPAEKTRVSEVLELPPHANHGNTFGGOIMAMENVAITLASFCAHPTLAKALEMFEHFRPGSDQNGRDLVLAUNNAFHGSMVEGICVAYVQAEETHRHINSAPFTFVLLDADDPQLLPWIRPFGDERRYRARSARKKIRIDRKITVSC

ORIGIN

Query Match 99.0%; Score 1839; DB 9; Length 6359;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QTEVPLSTVPMDBSNQVYLSYNNVSLKMLVADNWLSSRISGVRLTYLEDDKFLSFH
MEMVHVDAQAFLFLSLDRQREMDKHSRVELVQVVEDDAIYHTVSPALGHTKP
QDFVLIASRRKPCDNDGPVIALRSVLTLPHEHRETELCSQCLMEDEDOLTK
CCWRVSLTEIVSASGFYSWGLSRKRRSDGMNKLGGHSLTKLIPVAKINRFR
GYLDRT"

19 ATGATCCAGATGTCGAAATACCTCCGACCGGGGCTTGGCTCTGTGTTCTCCACCCG 78

Db 72 ATGATCCAGATGTCGAAATACCTCCGACCGGGGCTTGGCTCTGTGTTCTCCACCCG 131

QY 79 ACATCCCGGAAGTCAGCTTACGTCGCGGAGACAGACAGTCCATGAGCAGCGGAGGGA 138

Db 132 ACATCCCGGAAGTCAGCTTACGTCGCGGAGACAGACAGTCCATGAGCAGCGGAGGGA 191

QY 139 TACCGGAACCCCAACGAGGTGACAGATGAGCCAGCTGTGCTGCCACACCAACCA 198

Db 192 TACCGGAACCCCAACGAGGTGACAGATGAGCCAGCTGTGCTGCCACACCAACCA 251

QY 199 CGTGTGAGCTGAGCGCTCGGAGAGCTGCTCAAGTGTATACACACAGGCTTGCTGCC 258

Db 252 CGTGTGAGCTGAGCGCTCGGAGAGCTGCTCAAGTGTATACACACAGGCTTGCTGCC 311

QY 259 GCGGAGAGGACAGCTGAGCTCCCGCTGTGTACAGCTTCCATGATGATCATATTTAG 318

Db 312 GCGGAGAGGACAGCTGAGCTCCCGCTGTGTACAGCTTCCATGATGATCATATTTAG 371

QY 319 CACACCATTAATGTGAGCAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 378

Db 372 CACACCATTAATGTGAGCAAGTGTGATATATCAAGGCCAAGGTGAACCGGGCTTCAAC 431

QY 379 TCCAGATGAGAGTGTGGCATCCAGTGTGCTCCGAGAGACCTGTGCTGTGAAGCAATGG 438

Db 432 TCCAGATGAGAGTGTGGCATCCAGTGTGCTCCGAGAGACCTGTGCTGTGAAGCAATGG 491

QY 439 AATGTGCAAGGCTTGTGGCACCCTTGCTGTGGCCCGCGAGAGATCACCAAGTGAAGCTG 498

Db 492 AATGTGCAAGGCTTGTGGCACCCTTGCTGTGGCCCGCGAGAGATCACCAAGTGAAGCTG 551

QY 499 AAGCAGATCACCGCCGCGAGACAGAAAGAGAGATGAGACAGTGTGGCGGTGAGCGC 558

Db 552 AAGCAGATCACCGCCGCGAGACAGAAAGAGAGATGAGACAGTGTGGCGGTGAGCGC 611

QY 559 CGGCGCATGCGCTTGTGTCTATGACAGACCATCAAGACCTCTGTGGCACTGGCCATT 618

Db 612 CGGCGCATGCGCTTGTGTCTATGACAGACCATCAAGACCTCTGTGGCACTGGCCATT 671

QY 619 CAGGCGCATGTGAGAGACAGACTGTGAGCCAGTGTGCGGTGAGAGACCCGTGTG 678

Db 672 CAGGCGCATGTGAGAGACAGACTGTGAGCCAGTGTGCGGTGAGAGACCCGTGTG 731

QY 679 GAGAGTGTGAGAGTGTGCTGCTCCCTCCACGCAATCAACAGGCAACACTTTTGGGGC 738

Db 732 GAGAGTGTGAGAGTGTGCTGCTCCCTCCACGCAATCAACAGGCAACACTTTTGGGGC 791

QY 733 CAGATCATGCTGTGATGAGAAATGTGGCCACCATTCAGCCAGCCGGCTCTGCCGTGCC 798

Db 792 CAGATCATGCTGTGATGAGAAATGTGGCCACCATTCAGCCAGCCGGCTCTGCCGTGCC 851

QY 799 CACCTTAAGTGAAGGCAATGAAATGTTCACTTCGAGAGCCGCTCCAGGTGGAGAC 858

Db 852 CACCTTAAGTGAAGGCAATGAAATGTTCACTTCGAGAGCCGCTCCAGGTGGAGAC 911

QY 859 CGTCTGAGTCTCAAGCATGTGAACATATGCTTCAACATATGATGAGAGGTGGCGTG 918

Db 912 CGTCTGAGTCTCAAGCATGTGAACATATGCTTCAACATATGAGAGGTGGCGTG 971

QY 919 TGGTGAAGGCTTATGCGCAGAGGCTGAGACCCACCGGGCCACATCAACAGTGGCTTT 978

```

Db      972  TGGGTGAGGCGCTATCGCGAGGAGCTGAGACCCAGCGGCGGCACATGACAGTGCCTTT 1031
QY      979  ATGACCTTTGGTGTCTCTGAGGACATGACACGCCCGCTTGTCTGCTGGATTGGCC 1038
Db      1032  ATGACCTTTGGTGTCTCTGAGGACATGACACGCCCGCTTGTCTGCTGGATTGGCC 1091
QY      1039  CAGCCCGGCGATGTGAGCGGCGGTGACCGAGGCGGCGAGGCGGCGAGGCGGCGG 1098
Db      1092  CAGCCCGGCGATGTGAGCGGCGGTGACCGAGGCGGCGAGGCGGCGAGGCGGCGG 1151
QY      1099  GACAGGAGTACATGTGTCTCTGTAAGAGACAGAGTGTCTCTCTGCTGCTGCTG 1158
Db      1152  GACAGGAGTACATGTGTCTCTGTAAGAGACAGAGTGTCTCTCTGCTGCTGCTG 1211
QY      1159  CCTAGGACACAGGTGTCTCTGTAAGAGACAGAGTGTCTCTCTGCTGCTGCTG 1218
Db      1212  CCTAGGACACAGGTGTCTCTGTAAGAGACAGAGTGTCTCTCTGCTGCTGCTG 1271
QY      1219  AAGGACACTGGGTGTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCTG 1278
Db      1272  AAGGACACTGGGTGTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCTG 1331
QY      1279  GACAAAGTCTCTCTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCTG 1338
Db      1332  GACAAAGTCTCTCTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCTG 1391
QY      1339  CTGCTGCTCTCTCTCTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCT 1398
Db      1392  CTGCTGCTCTCTCTCTCTCTCTGAGAGTGTCTCTGAGTGTCTCTGAGTGTCT 1451
QY      1399  CTAGTGTGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1458
Db      1452  CTAGTGTGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1511
QY      1459  GGTACACAAAGCCCGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1518
Db      1512  GGTACACAAAGCCCGAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1571
QY      1519  GGGGACCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1578
Db      1572  GGGGACCCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1631
QY      1579  GAGTACAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
Db      1632  GAGTACAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
QY      1639  CTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1698
Db      1692  CTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1751
QY      1699  TATTCCTGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1758
Db      1752  TATTCCTGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1811
QY      1759  CTAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818
Db      1812  CTAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1871
QY      1819  TTTGGATACCTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Db      1872  TTTGGATACCTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1910

```

```

RESULT 3
AP416922 1818 bp mRNA linear PRI 13-DEC-2001
LOCUS Homo sapiens brown fat inducible thioesterase 2 mRNA, complete cds,
DEFINITION alternatively spliced.
ACCESSION AP416922
VERSION AP416922.1 GI:17646238
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 1818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
23. 1807
/feature="BFI2; acyl-CoA thioesterase; acyl-CoA hydrolase;
alternatively spliced"
/codon_start=1
/product="brown fat inducible thioesterase 2"
/protein_id="AA140938.1"
/db_xref="GI:17646239"
/translation="MIQWNGNHLRRGLASVFSNRTSRKSLRAGNDSAMADGEGYRNP
TEVMSQVILPECHTNRGELSGQLLKIITPACISARHAGPCVTSMDIYEHT
ISGVQVNIKAKNRAFSNMEVGVQVSELCISKONKALATPARSEBITVKT
KOIPIREEMKESHSVAERBRMLVYADTLKDLANALIGDLESRSNMPAREKT
RVSEVELVPRPHANGTNPFGQIMAMENAVTIAASLCANPILKAIEMFHPGRS
QVGRILYKALVNAFHSMEVGVCEAVRGEAEHRHINSAPFTFVLADDPOL
LPWIRPQDGERRYREASARKKRLDRKYIVSCQTEVPLSVDPENQVLLSDYRN
ESLKLVAKDMWVLSSEISQVRLTLEBDELSPHEMNVHDAQAQALISDRORP
BMDHYSVELVOOVEDDALYHTSPALGGHTKPODVLILASRRKPCDNDPVLAL
RSVTLPHREPERYREGTICSGPICMEBGOILTQVSYNNAATREVLNVYTTNAGLS
SEFTTTRACQFLLDRNDLAPSLQTL"

```

ORIGIN

```

Query Match      88.0%; Score 1633.6; DB 9; Length 1818;
Best Local Similarity 99.8%; Pred. No. 4.9e-308;
Matches 1636; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

15  TCAATGATCCAGATGTGCGAAATCACTGCGAGCGGGCTTGCTGTGTTCTCCA 74
Db      19  TCAATGATCCAGATGTGCGAAATCACTGCGAGCGGGCTTGCTGTGTTCTCCA 78
QY      75  CCGACATCCCGGAAGTACAGCTTACGTCGGGGGAGACAGAGTGCATGGCGAGCG 134
Db      79  CCGACATCCCGGAAGTACAGCTTACGTCGGGGGAGACAGAGTGCATGGCGAGCG 138
QY      135  GGGATACCGGAACCCACAGAGGTGTCAGATGAGCCAGTGTGCTGCCCTGCCACCA 194
Db      139  GGGATACCGGAACCCACAGAGGTGTCAGATGAGCCAGTGTGCTGCCCTGCCACCA 198
QY      195  CCAACGTGTGAGCTGAGCGTGTGGGACAGCTGCTCAAGTGTGATGACACACAGG 254
Db      199  CCAACGTGTGAGCTGAGCGTGTGGGACAGCTGCTCAAGTGTGATGACACACAGG 258
QY      255  GTCCGCGGAGAGGACAGCTGTGCGCCCTGTGTGACACAGCTTCCATGATGATATT 314
Db      259  GTCCGCGGAGAGGACAGCTGTGCGCCCTGTGTGACACAGCTTCCATGATGATATT 318
QY      315  TGAGACACCATTAAGTGTGAGCAAGTGTGATATCAAGGCCAAGGTGAACCGGCG 374
Db      319  TGAGACACCATTAAGTGTGAGCAAGTGTGATATCAAGGCCAAGGTGAACCGGCG 378
QY      375  CAATCCAGCATGAGAGTGTGGGATTCAGAGTGTGCTGCGAGGACCTGTGCTGAGA 434
Db      379  CAATCCAGCATGAGAGTGTGGGATTCAGAGTGTGCTGCGAGGACCTGTGCTGAGA 438

```

[illegible]

Db	1519	CAATGGGAGCCCTATGTATCGCGCTGAGAGTGGGTCAACCTGCGCACACACCGAGAGAC	1578
Qy	1575	GCCAGAGTACAGACGCCGAGAGACCTCTGTGCTAAGGCTTCTGCTCTGGGCGCAGGGGGGA	1634
Db	1579	GCCAGAGTACAGACGCCGAGAGACCTCTGTGCTAAGGCTTCTGCTCTGGGCGCAGGGGGGA	1638
Qy	1635	CCAGCTGACCAAGTCGCT	1654
Db	1639	CCAGCTGACCAAGTATCTT	1658
RESULT 4			
AX883384			
LOCUS	AX883384	3082 bp	DNA linear PAT 17-DEC-2003
DEFINITION	Sequence 18289 from Patent EP1074617.		
ACCESSION	AX883384		
VERSION	AX883384.1	GI:40038285	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
JOURNAL	Primers for synthesizing full-length cDNA and their use		
FEATURES	Patent: EP 1074617-A 18289 07-FEB-2001; Research Association for Biotechnology (JP)		
SOURCE	location/Qualifiers		
CDS	1..3082		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	83..1867		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAB91952.1"		
	/db_xref="GI:40038286"		
	/translation="MIQNVGNLHRRGLASVPSNRTSRKSLRAGNDSAMADGEGYRNFTEVMSQVLPECHTNGELSVQLKMWIDTTCILAEHRHGCCEVTASMDIYFEHTISVGVNVIKAKVRANFNSMEVGIQVASELCEKQNVCKALATFYAREIRTEVKIKQIPIREBEKMSHVAERRMRLIVADDTKLIANALQVEDSPRCSRMVPAEKIKVSEVLELPEPHAHQGTNRGGQIMAMKENVATTAASRLCAHPTTLKAIEMFHPGPEQVGRVLVKAIVNNAFHSMEVGVCEVAERAEHRHINSAPMTFVVLADADPOQLPMPRPQDGEERYERYEASARKIRLDRKYIVSOKTEVPSVMDPSNOYLSNNYSLSMLVAKNWVLSSEISQVRLYTEIDDKFLSPHEMENVVDAQAQFLISDRORRSMDSHRSVELVOGVDEDDATVHTSPALGGTROPDVLIASRRKPCDNBDPYIALRSYVLPHRETRPREYRRGETLCSGFLWRMSGQULKVSYNNAATGELVNYVTNVAGLSSEFYTFRACQEPFLDNBNDLAPLSQTL"		
ORIGIN			
Query Match	87.7%: Score 1628; DB 6; Length 3082;		
Best local similarity	99.7%: Pred. No. 5,7e-307;		
Matches 1631; Conservative	0; Mismatches 5; Indels 0; Gaps 0		
Qy	19	ATGATCCAGAAATGTGGAATACCTGCGACGGGGCTTGCCCTGTGTCTTCCAAACGC	78
Db	83	ATGATCCAGAAATGTGGAATACCTGCGACGGGGCTTGCCCTGTGTCTTCCAAACGC	142
Qy	79	ACATCCCGGAAGTCAGCTTACGTGGGGGGAACGACATGCCATGGACAGCGGACAGGA	138
Db	143	ACATCCCGGAAGTCAGCTTACGTGGGGGGAACGACATGCCATGGACAGCGGACAGGA	202
Qy	139	TACCGGAACCCCAACGAGGTGCGAGATGAGACAGCTGGTGCTGCCCTGCCACCAACCA	198
Db	203	TACCGGAACCCCAACGAGGTGCGAGATGAGACAGCTGGTGCTGCCCTGCCACCAACCA	262
Qy	199	CGTGGTAGCTGAGCGGTGGGAGCTGTGCTCAAGTGATTTAGACCAACGAGCTTGTCTTC	258
Db	263	CGTGGTAGCTGAGCGGTGGGAGCTGTGCTCAAGTGATTTAGACCAACGAGCTTGTCTTC	322
Qy	259	GCGAGAGACACGCTGGTCCCTCTGTGTACACGTTCCATGAGATGACATCATTTTAG	318

```

|||
323 GCGAGAGGCAAGCTGGTGGTCCCTGTGTCAAGCTTCATGATGACATCTATTTTGAG 382
QY 319 CACACCATTAAGTGTGGACAAGTGGTAATATCAAGGCAAGGTGAACCGGGCTTCAAC 378
Db 383 CACACCATTAAGTGTGGACAAGTGGTAATATCAAGGCAAGGTGAACCGGGCTTCAAC 442
QY 379 TCCAGCATGAGTGGGACATCCAGGTGGCTGGAGGACCTGTGCTCTGAGAAAGCATG 438
Db 443 TCCAGCATGAGTGGGACATCCAGGTGGCTGGAGGACCTGTGCTCTGAGAAAGCATG 502
QY 439 AATGTGTGAAGGCTTGGCCACCTTCTGTGGCCCGCGAGAGATCAACCAAGTGAAGCTG 498
Db 503 AATGTGTGAAGGCTTGGCCACCTTCTGTGGCCCGCGAGAGATCAACCAAGTGAAGCTG 562
QY 499 AAGCAGATACAGCCCGCGAGACAGAAAGAGGAAGATGAGACAAGTGTGGCGCTGAGCGC 558
Db 553 AAGCAGATACAGCCCGCGAGAGCAAGAAAGAGGAAGATGAGACAAGTGTGGCGCTGAGCGC 622
QY 559 CGGCGCATGCGCTGTGTATGACAGACATCAAGGACCTCTGGCCCAACTGGCCCAT 618
Db 623 CGGCGCATGCGCTGTGTATGACAGACATCAAGGACCTCTGGCCCAACTGGCCCAT 682
QY 619 CAGGCGCATCTGGAGAGCAGAGACTGTAGCCGATGTGCTGAGAGAGACCGCTGTG 678
Db 683 CAGGCGCATCTGGAGAGCAGAGACTGTAGCCGATGTGCTGAGAGAGACCGCTGTG 742
QY 679 GAGAGTGTGAGCTGTGCTGCTGCTCCCAAGCAATCAACAGGCGCAACCTTTGGGGGC 738
Db 743 GAGAGTGTGAGCTGTGCTGCTGCTCCCAAGCAATCAACAGGCGCAACCTTTGGGGGC 802
QY 739 CAGATCATGCGCTGTGATGAGAAATGTGGCACTTGGACGCGGCTGCGCTGCG 798
Db 803 CAGATCATGCGCTGTGATGAGAAATGTGGCACTTGGACGCGGCTGCGCTGCG 862
QY 799 CACCTTACGCTGAAGGCCATTGAATGTTCCATTCCAGAGCCGCTCCCAAGTCCGCGAC 858
Db 863 CACCTTACGCTGAAGGCCATTGAATGTTCCATTCCAGAGCCGCTCCCAAGTCCGCGAC 922
QY 859 CGTGTGTGTCTCAAGGCCATGTGAAACATGCTTCAAAACATAGCATGAGTGGCGCTG 918
Db 923 CGTGTGTGTCTCAAGGCCATGTGAAACATGCTTCAAAACATAGCATGAGTGGCGCTG 982
QY 919 TGGGTGAGAGGCTTATCGCCAGAGGCTGAGACCCAGCGCGCAATCAACAGTGGCTT 978
Db 983 TGGGTGAGAGGCTTATCGCCAGAGGCTGAGACCCAGCGCGCAATCAACAGTGGCTT 1042
QY 979 ATGACCTTTGTGTCTGAGACGCAAGATGACAGCCCAAGTGTGCTGATTCGAGCC 1038
Db 1043 ATGACCTTTGTGTCTGAGACGCAAGATGACAGCCCAAGTGTGCTGATTCGAGCC 1102
QY 1039 CAGCCCGCGCATGAGAGCGCGCGATGACCGAGAGGCAATGCGCAAAAAGATCCGCTG 1098
Db 1103 CAGCCCGCGCATGAGAGCGCGCGATGACCGAGAGGCAATGCGCAAAAAGATCCGCTG 1162
QY 1099 GACAGGAAGTACATGCTGTCTGTAAAGACAGAGTGGCCCTCTGCTCCCTGGAGC 1158
Db 1163 GACAGGAAGTACATGCTGTCTGTAAAGACAGAGTGGCCCTCTGCTCCCTGGAGC 1222
QY 1159 CTTAGCAACAGGTGTACTGAGCTAACATAACGTCTCTCTTGAAGATCTTGTGCGC 1218
Db 1223 CTTAGCAACAGGTGTACTGAGCTAACATAACGTCTCTCTTGAAGATCTTGTGCGC 1282
QY 1219 AAGAGCAACTGGGTGTCTCTCTCGAGATCACTAGGTCGCGCTGTACTCTGAGAT 1278
Db 1283 AAGAGCAACTGGGTGTCTCTCTCGAGATCACTAGGTCGCGCTGTACTCTGAGAT 1342
QY 1279 GACAAGTTCCTCTCTCTCAATGAGATGAGTGGTGCATGTGAGTGCAGCCAGGCTTC 1338
Db 1343 GACAAGTTCCTCTCTCTCAATGAGATGAGTGGTGCATGTGAGTGCAGCCAGGCTTC 1402
QY 1339 CTGCTGCTCTGAGCTGTGCTGAGAGGCGAGAGTGGGCAAGCACTACCGAGGCTGAG 1398

```

```

Db 1403 CTGCTGCTCTGGAACCTGCGCTCAGAGGCTCAGAGTGGGACACACTACCGAGGCTGGAG 1462
QY 1399 CTAGTGCACAGTGTAGACGAGAGCAGACGCGCATCTACCAAGTCAAGAGCCCTGCTGGA 1458
Db 1463 CTAGTGCACAGTGTAGACGAGAGCAGACGCGCATCTACCAAGTCAAGAGCCCTGCTGGA 1522
QY 1459 GGTCAACAAAGCCCGAGACTTGTGATCTCTGAGTCTGAGGCGGAAGCTTGTGCAAT 1518
Db 1523 GGTCAACAAAGCCCGAGACTTGTGATCTCTGAGTCTGAGGCGGAAGCTTGTGCAAT 1582
QY 1519 GGGGACCCCTATGTATGATCGCGCTGAGAGTGTGATCGCTGCGCCACACAGAGAGAGCCCA 1578
Db 1583 GGGGACCCCTATGTATGATCGCGCTGAGAGTGTGATCGCTGCGCCACACAGAGAGAGCCCA 1642
QY 1579 GAGTACAGACGCGGAGAGACCTTGTGCTGAGGCTTGTGCTGCGCGAGGAGGAGCCAG 1638
Db 1643 GAGTACAGACGCGGAGAGACCTTGTGCTGAGGCTTGTGCTGCGCGAGGAGGAGCCAG 1702
QY 1639 CTGACCAAGTGTGCT 1654
Db 1703 CTGACCAAGTGTGCT 1718

RESULT 5
BD160291
LOCUS BD160291
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160291
VERSION BD160291.1 GI:27866049
KEYWORDS JP 2002191363-A/15134.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 3082)
REFERENCE
Ota,T., Iwagui,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 15134 09-JUN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/15134
PD 09-JUN-2002
PF 28-JUN-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KOORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(83)..(1864).
FT CDS
Location/Qualifiers
1..3082
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 87.7%; Score 1628; DB 6; Length 3082;
Best Local Similarity 99.7%; Pred.No. 5.7e-307;
Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATGCGGAAATCACCTGCGAGCGGGCTTGGCTCTGTCTTCCAAAGCC 78
Db 83 ATGATCCAGAAATGCGGAAATCACCTGCGAGCGGGCTTGGCTCTGTCTTCCAAAGCC 142
QY 79 ACATCCGGAAGTACAGCTTACGTGCGGGAAGCAGACAGTGCATAGGCGAGCGGAGGGA 138
Db 143 ACATCCGGAAGTACAGCTTACGTGCGGGAAGCAGACAGTGCATAGGCGAGCGGAGGGA 202

```


CDS

```

/clone.lib="THYR01"
/note="Cloning vector: pME18SFL3"
63..1867
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14734.1"
/db_xref="gi:10436032"
/translation="MIQVGNHLRGLASVPSNRTSRKSLRAGNDSAMADGEYRNP
TEVMSOLVLPCHTMRGELISVGLLKIPIIDTACLSARHAGCPVTASMDITFHT
ISVGVNVIKAVNRARNSMEVGIQVASEDISKQNVKALTFVARREITVKL
KQITPRIEEKMEHSVAERRRRLVADTLKDLANCAIQDELRDSCSRVPEKT
QVESVELLPVPHNQNTFGQIMAMNENATIIASRLIASHAPILKAIEMEPKPS
QVDELIVKAIYVNAFKHSMVEYGVCEAVRQAEHRHISAFTFVVLADDDPOL
LPIWIPQDGERRYREASARKILDKYIVSCOTPEVLPSPMDPNQVYLSTNNY
SLKMLVAKDNWVSSSEISQVRLYTLDEDKELSPMEWVYVDAQAFLISLDFORP
EMDKHRSVELVQVDEDDAIYHTYSPALGHTKPODVIILASRKPCKDNEDPVIAL
RSVTLPIHRETPHREGLCSGFLMEBQDLTKVSYNNQATPEVLNVYTTNAGLS
SEFYTFRACBQFLIDNRNDLAPSLQTL"

```

ORIGIN

Query Match 87.7% Score 1628; DB 9; Length 3082;
 Best Local Similarity 99.7% Pred. No. 5,7e-307;

Matches 1631; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

19 ATGATCCAGAAATGTCGAAATCCTGCGACGCGGCTTG3CCTGTGTCTCCAAACGC 78
83 ATGATCCAGAAATGTCGAAATCCTGCGACGCGGCTTG3CCTGTGTCTCCAAACGC 142
79 ACATCCCGGAAAGTCAGCTTACGTGCGGAGAAAGCAAGTCCCATGCGACGCGAG3A 138
143 ACATCCCGGAAAGTCAGCTTACGTGCGGAGAAAGCAAGTCCCATGCGACGCGAG3A 202
139 TACCGGAAACCCGACGAGGATGAGTGAAGCAGCTGCTGCTGCGACCAACCAACAA 198
203 TACCGGAAACCCGACGAGGATGAGTGAAGCAGCTGCTGCTGCGACCAACCAACAA 262
199 CGTGTGAGCTGAGAGCGTGGGAGCTGCTCAAGTGAATGACACCAACGCTTGCTCC 258
263 CGTGTGAGCTGAGAGCGTGGGAGCTGCTCAAGTGAATGACACCAACGCTTGCTCC 322
259 GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
323 GCGGAGAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 382
319 CACACCATTAAGTGTGACAGAGTGTGAATTAAGGCAAGGCAAGGCAAGGCAAGGCA 378
383 CACACCATTAAGTGTGACAGAGTGTGAATTAAGGCAAGGCAAGGCAAGGCAAGGCA 442
379 TCCAGCATGAGAGTGGGATCCAGGTGCTGCGAGGACCTGTGCTGAGAGAGAGTGG 438
443 TCCAGCATGAGAGTGGGATCCAGGTGCTGCGAGGACCTGTGCTGAGAGAGAGTGG 502
439 AATGTGTGCAAGGCTTGGCCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 498
503 AATGTGTGCAAGGCTTGGCCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562
499 AAGCAGATCAAGCGCGGACAGAGAGAGAGAGATGAGCAAGGTGCGCGCTGAGCGC 558
563 AAGCAGATCAAGCGCGGACAGAGAGAGAGAGATGAGCAAGGTGCGCGCTGAGCGC 622
559 CGGCGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTGCGCAATGCGCCAT 618
623 CGGCGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTGCGCAATGCGCCAT 682
619 CAGGCGCATCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 678
683 CAGGCGCATCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 742
679 GAGAGTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 738
743 GAGAGTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802
739 CAGATCATGCGCTTGATGAGAGATGTGAGCAACATGAGCGACGCGCTGCTGCTGCT 798

```

```

Db 803 CAGATCATGCGCTTGATGAGAGATGTGAGCAACATTTGACAGCGCGCTGCGGTGCC 862
799 CACCTTAAGCTTGAAGGCAATTTAAATGTTTCACTTCCAGAGCGCGTCCAGAGTCCGAGC 858
863 CACCTTAAGCTTGAAGGCAATTTAAATGTTTCACTTCCAGAGCGCGTCCAGAGTCCGAGC 922
859 CGTCTGTGTCTTAAAGGCAATGTTGAACATGCTTCAACATATGATGAGAGTGGCGTGG 918
923 CGTCTGTGTCTTAAAGGCAATGTTGAACATGCTTCAACATATGATGAGAGTGGCGTGG 982
919 TCGGAGAGGCGCTTATGCGCAGAGAGCTGAGACCCACCGCGCGCAATCAAGAGTGGCTT 978
983 TCGGAGAGGCGCTTATGCGCAGAGAGCTGAGACCCACCGCGCGCAATCAAGAGTGGCTT 1042
979 ATGACCTTGTGTGCTTGGAGCGAGATGACAGCGCCCAAGTTGCTGCTGATTTGGGCC 1038
1043 ATGACCTTGTGTGCTTGGAGCGAGATGACAGCGCCCAAGTTGCTGCTGATTTGGGCC 1102
1039 CAGCCCGCGAGTGTGAGCGCGGCTTACCGAGAGGCGCAATGCGCAAAAGATCCGCTTG 1098
1103 CAGCCCGCGAGTGTGAGCGCGGCTTACCGAGAGGCGCAATGCGCAAAAGATCCGCTTG 1162
1099 GACAGGAAGTACATGCTGCTGTAAGCAGACAGAGTGCCTCTGCTGCTGCTGCTGCTG 1158
1163 GACAGGAAGTACATGCTGCTGTAAGCAGACAGAGTGCCTCTGCTGCTGCTGCTGCTG 1222
1159 CCTAGCAACAGGTGTACCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1218
1223 CCTAGCAACAGGTGTACCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1282
1219 AAGGCAACAGGTGTACCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1278
1283 AAGGCAACAGGTGTACCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1342
1279 GACAAAGTCTCTCTCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1338
1343 GACAAAGTCTCTCTCTTGAAGCTCAATATAAGTCTCTCTCTTGAAGTCTTGGCC 1402
1339 CTGCTGCTCTTGAAGCTTGTGAGAGCGCGAGAGTGGGCAAGACTTGAAGTGGAG 1398
1403 CTGCTGCTCTTGAAGCTTGTGAGAGCGCGAGAGTGGGCAAGACTTGAAGTGGAG 1462
1399 CTAGTGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
1463 CTAGTGACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1522
1459 GGTACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
1523 GGTACACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582
1519 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
1583 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1642
1579 GAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
1643 GAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1702
1639 CTGACCAAGTGTGCT 1654
1703 CTGACCAAGTGTGCT 1718

```

RESULT 7
 AK091112
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ33793 fis, clone CD34C200043, weakly similar
 to Aey1-coA hydrolase.
 ACCESSION AK091112
 VERSION AK091112.1 GI:21749405
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1
Ninomiya, K., Magatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, B., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Mateno, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kaneshiro, K., Takahashi-Fujii, A., Ohshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2875)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' & 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES
source
1. 2875
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CD34C2000443"
/cell_type="CD34+ Cells"
/note="cloning vector: pME18SFL3-primary culture, CD34+ Cells"
ORIGIN
Query Match 78.1%; Score 1450; DB 9; Length 2875;
Best Local Similarity 95.0%; Pred. No. 2,7e-272;
Matches 1536; Conservative 0; Mismatches 10; Indels 70; Gaps 1;
QY 39 TCACCTCGGAGCGGGCTTGCGCTGTGTTCTCCACCGGACATCCCGGAAGTACGCTT 98
DB 169 TCCCATCTTTGGGGCTTGGCTCTGTGTCTCCACCGGACATCCCGGAAGTACGCTT 228
QY 99 ACGTGCAGGGAACACAGTGCATGCGAGACGCGAGGAGATCCCGAACCACGAGGT 158
DB 229 ACGTGCAGGGAACACAGTGCATGCGAGACGCGAGGAGATCCCGAACCACGAGGT 288
QY 159 GCAGATAGCCAGCTGTGTCTGCTCCCTGCGACACCAACCACTGTGTAGCTGAGCGTGG 218
DB 289 GCAGATAGCCAGCTGTGTCTGCTCCCTGCGACACCAACCACTGTGTAGCTGAGCGTGG 348
QY 219 GCAGCTCTCAAGTGAATTGACACCGCGCTTGGCTGCGCGGAGAGGACGCTGGGTG 278
DB 349 GCAGCTCTCAAGTGAATTGACACCGCGCTTGGCTGCGCGGAGAGGACGCTGGGTG 388
QY 279 CCCCTGTGTACAGCTTCCATGATGACATCTATTGTGAGCACACATTAGTGTGGACA 338
DB 389 -----GTGTTGAGCA 398
QY 339 AGTGTGAATATCAAGGCCAAAGTGAACCGGGCTTCAATCTCGAGCATGAGGTGGCAT 398
DB 399 AGTGTGAATATCAAGGCCAAAGTGAACCGGGCTTCAATCTCGAGCATGAGGTGGCAT 458
QY 399 CCAAGTGGCTTCGAGAGCACTGTGCTCGAAGAGAGGAAATGTGTGCAAGGCTTGGC 458
DB 459 CCAAGTGGCTTCGAGAGCACTGTGCTCGAAGAGAGGAAATGTGTGCAAGGCTTGGC 518

QY 459 CACCTTCGTGGCCCCCGGAGAGATCAACCAAGTGAAGCTGAAGACATACCGCCGGGAC 518
DB 519 CACCTTCGTGGCCCCCGGAGAGATCAACCAAGTGAAGCTGAAGACATACCGCCGGGAC 578
QY 519 AGAAGAGGAGAGATGAGAGACAGTGTGGCGGTGAAGCGCGCGCATGGCGCTTGTCTA 578
DB 579 AGAAGAGGAGAGATGAGAGACAGTGTGGCGGTGAAGCGCGCGCATGGCGCTTGTCTA 638
QY 579 TCGAGACACCATCAAGGACCTCTCTGGCCAACTGCGCAATTCAGGGCGATCTGAGAGCAG 638
DB 639 TCGAGACACCATCAAGGACCTCTCTGGCCAACTGCGCAATTCAGGGCGATCTGAGAGCAG 698
QY 639 AGACTGTAGCGCATGAGTGGCTGGCTGAGAGACCCGTGTGAGAGTGTGAGAGCTGTCT 758
DB 699 AGACTGTAGCGCATGAGTGGCTGGCTGAGAGACCCGTGTGAGAGTGTGAGAGCTGTCT 758
QY 699 GCGTCCCGGAGGCAATTCACAGGGGCAACACCTTTGGGGGCGAGATCAATGGCTGGATGGA 758
DB 759 GCGTCCCGGAGGCAATTCACAGGGGCAACACCTTTGGGGGCGAGATCAATGGCTGGATGGA 818
QY 759 GAATGTGGCCACCATTCGACAGCGGCTCTGCGCTGCGCCACCTTACGCTGAGAGGCGAT 818
DB 819 GAATGTGGCCACCATTCGACAGCGGCTCTGCGCTGCGCCACCTTACGCTGAGAGGCGAT 878
QY 819 TGAATGTTCACCTTCGAGGCGCGTCCAGGTGTGCGGACCGCTTGTGCTCAAGCGCAT 878
DB 879 TGAATGTTCACCTTCGAGGCGCGTCCAGGTGTGCGGACCGCTTGTGCTCAAGCGCAT 938
QY 879 CGTGAACAAATGCGCTTCAAAATGACATGAGAGGTGGGGGTGTGGGTGAGGCGCTATCGCA 938
DB 939 CGTGAACAAATGCGCTTCAAAATGACATGAGAGGTGGGGGTGTGGGTGAGGCGCTATCGCA 998
QY 939 GGAAGCTGAGACCCACCGGCGGCAATCAACAGTGCCTTTATGACCTTTGTGTCTTGA 998
DB 999 GGAAGCTGAGACCCACCGGCGGCAATCAACAGTGCCTTTATGACCTTTGTGTCTTGA 1058
QY 999 CGAGATGACACGAGCGGAGTTGCTGCGCTGAGATTGGCGCCGAGCGGAGTGTGAGCG 1058
DB 1059 CGAGATGACACGAGCGGAGTTGCTGCGCTGAGATTGGCGCCGAGCGGAGTGTGAGCG 1118
QY 1059 GCGGTACCGAGAGCGCGGCGGCAATCAACAGTGCCTTTATGACCTTTGTGTCTTGA 1118
DB 1119 GCGGTACCGAGAGCGCGGCGGCAATCAACAGTGCCTTTATGACCTTTGTGTCTTGA 1178
QY 1119 CTGTAAACAGACAGAGGTGCGCTCTCTGCGTCCCTGCGGACCCCTGAGCAACGAGTACT 1238
DB 1179 GAGCTAACATPAAGTCTCTCTCTTGAAGATGCTTGTGGCCAAAGACAACTGGTGTCTC 1238
QY 1179 GAGCTAACATPAAGTCTCTCTCTTGAAGATGCTTGTGGCCAAAGACAACTGGTGTCTC 1298
DB 1239 GAGCTAACATPAAGTCTCTCTCTTGAAGATGCTTGTGGCCAAAGACAACTGGTGTCTC 1298
QY 1239 CTGGAGATCAGTACAGTCCGCGCTGACCTCTGAGAGATGACAAAGTTCCTCTCTTCCA 1298
DB 1299 CTGGAGATCAGTACAGTCCGCGCTGACCTCTGAGAGATGACAAAGTTCCTCTCTTCCA 1358
QY 1299 CATGAGATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1358
DB 1359 CATGAGATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418
QY 1359 TCAGAGGCGAGAGTGGGACAAAGCATACCGAGAGGTGAGATGAGAGCGAGTGAAGGA 1418
DB 1419 TCAGAGGCGAGAGTGGGACAAAGCATACCGAGAGGTGAGATGAGAGCGAGTGAAGGA 1478
QY 1419 GGAAGAGCGCATTAACAGCGTCAACAGCGCGCTCGCTGGAGAGTCAACAAAGCCCGAGA 1478
DB 1479 GGAAGAGCGCATTAACAGCGTCAACAGCGCGCTCGCTGGAGAGTCAACAAAGCCCGAGA 1538
QY 1479 CTTGTGATCTGTGCTGAGGCGGAGGCTTGTGACATGAGGAGCCCTTATGTATGCTGC 1538
DB 1539 CTTGTGATCTGTGCTGAGGCGGAGGCTTGTGACATGAGGAGCCCTTATGTATGCTGC 1598
QY 1539 GCTGAGGTGCTGACGCTGCGCCACACACCGAGAGACGCGCAGAGTACAGACGCGGAGAGAC 1598

Db 1599 GGTGAGTGTGCTGACGCTGCCCCACACCGAGAGCGCAGAGTACAGACCGGAGAGAC 1658
QY 1599 CCTGCTCAGGCTTCTGCTCTGCGCGAGAGGAGACCGCTGACCAAGTCTGCT 1654
Db 1659 CCTGCTCAGGCTTCTGCTCTGCGCGAGAGGAGACCGAGTACCAAGTATCTCT 1714

RESULT 8
AF416923 2678 bp mRNA linear ROD 13-DEC-2001
LOCUS AF416923
DEFINITION Mus musculus brown fat inducible thioesterase 2 mRNA, complete cds.
ACCESSION AF416923
VERSION AF416923.1 GI:17646240
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2678)
Adams, S.H., Chui, C., Schilbach, S.L., Yu, X.X., Goddard, A.D., Grimaldi, C., Lee, J., Dowd, P., Colman, S. and Lewin, D.A.,
Bf17, a unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adipose Tissue. Cloning, Organization of the Human Gene, & Assessment of a Potential Link to Obesity
Biochem. J. (2002) In press
2 (bases 1 to 2678)
Adams, S.H., Chui, C., Schilbach, S.L., Yu, X.X., Goddard, A.D., Grimaldi, C., Lee, J., Dowd, P., Colman, S. and Lewin, D.A.,
Direct Submission
Submitted (07-SEP-2001) Endocrinology, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA

FEATURES
SOURCE
1..2678
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/chromosome="4"
54..1838
/note="Bf172; acyl-CoA thioesterase; acyl-CoA hydrolase; similar to Homo sapiens Bf172"
/product="Brown fat inducible thioesterase 2"
/codon_start=1
/protein_id="AA14093.1"
/db_xref="gi:17646241"

CDs
/translation="MIONVGNHLRGRFASMSNRTSRKSIHPESGDPPTMAEBGR NPTVQMSQVLPCHNHRGELISGLKWIIDTACLSAERHACPCVTASMDIYFD HTISVQGVNLIKAVNRAFNSMEVGIQVSEIDICSEKGMVACALAFVHRELISKY KLRQVILPIIEEKTEHGVAAERRMRLVYADTIKDLTLHCVIDLDDKCSNMPAEK TRVESVILVEPHNHOENTFGQIMAMEVATIAASRLCHAHPTLKAIMPHRGP SOWGDRVLKATVNNAFKSMVEGVCEVAQOEATROPHNNSAPMPFVILDKDQO KLPMPRQPGEGRRRARRKRIKRIIDRKIVSCQKAEVALSYMPDPSNOVITYSYN VSLIKTLAKDNWLVSEIVRLYLIEEDFLSHLEWVNVNDAQVQLSDIRRP EMDHYRSEVLEQVDEDDAIYHVISPALSENTPKDFVILASRRKCDNDPVIAR RSVTLPIHNETPEYQREGTLCSGFLMRBEGQMTKVSYNQATPGFIYVTTNVSGLS SEFYNTKACSFLLDNRNLDAPSLQTL"

ORIGIN
Query Match 66.2%; Score 1230; DB 10; Length 2678;
Best Local Similarity 85.8%; Pred. No. 1.9e-229;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;

QY 18 AATGATCCAGAAATGTCGGAATCACTGCGAGCGGGGCTTGCTGTTCTTCCACCG 77
Db 53 AATGATCCAGAAATGTCGGAATCACTGCGAGCGGGGCTTGCTGTTCTTCCACCG 112
QY 78 CACATCCCGAAGTCAAGCTTACGTGCG-----GGGAACGACAGTGCATGAGACGG 131
Db 113 CACATCCCGAAGTCAAGCTTCCCATCGAGGCTCGAGAGACCTCTACATGAGAGGG 172
QY 132 CGAGGATACCGGACCCGACGAGGTGACATAGCCAGTGTGCTGCCCTGCCACAC 191

Db 173 TGAAGATACCGGAACCCACGAGAGGTGAGATGAGCCAGCTGTACTGCCCTGCCACAC 232
QY 132 CAACCAAGCTGTGAGCTGAGCGTGGGCAAGTCTCAAAGTGAATTGACACACGCGTTG 251
Db 233 CAACCAAGCTGTGAGCGTGGGCAAGTCTCAAAGTGAATTGACACACGCGTTG 292
QY 252 CCTGTCCGAGAGAGACAGCTGCGCGCCCTGTGTCAACAGCTTCCATGAGATGACATCTA 311
Db 293 CTTATCAGGAGAGAGAGAGCTGTGCTGTCTCCCTGACAGCTTATGATGATGACATCTA 352
QY 312 TTTTGAACACACCATTAAGTGTGGAACAAGTGTGAATATCAAGCCAAAGTGAACCGGCG 371
Db 353 CTTGACCATATTCATTAAGTGTGCGCAAGTGTGTGATATCAAGCCAAAGTGAACCGGCG 412
QY 372 CTTCAATTCACGATGAGAGTGGGCAATCCAGGTGCTTGGAGAGACCTGTGCTTGAGAA 431
Db 433 CTTCAATTCACGATGAGAGTGGGCAATCCAGGTGCTTGTGAGAGATCTGTGCTTGAGAA 472
QY 432 GCAAGTGAATGTGTGCAAGAGCTTGGCCACCTTGTGTGCGCGCGAGAGATGACCAAGGT 491
Db 473 GCAAGTGAATGTGTGCAAGAGCTTGGCCACCTTGTGTGCGCGCGAGAGATGACCAAGGT 532
QY 492 GAAAGTGAAGCAGATCACGCGCGAGACAGAGAGAGAGAGATGAGACACAGTGTGGCGCG 551
Db 533 GAAAGTGAAGCAGATCACGCGCGAGACAGAGAGAGAGAGATGAGACACAGTGTGGCGCG 592
QY 552 TGAAGCGCGGCGATGCGGCTTGTGTATGACAGACACCATCAAGACCTTCTGGCCACTG 611
Db 593 TGAAGCGCGGCGATGCGGCTTGTGTATGACAGACACCATCAAGATCTCTTAACCCACTG 652
QY 612 CGCCATTCAGGCGCATGTGAGAGACAGACCTGTGCGCGCATGTGTGCGCGCGAGAGAC 671
Db 653 TGTATCATCAGACCATTTTGA---CAAGACCTGACAGCATATGTGTGCGCGAGAGAC 709
QY 672 CGGTGTGAGAGTGTGAGAGCTGTCTGCTCTCCCAACGCAATCAACGAGGCAACACTT 731
Db 710 CCGAGTGAAGAGTGTGAGAGCTGTGTCTCTCTCCCAACGCAATCAACGAGGCAATCACTT 769
QY 732 TGGGGGCGCATCATGCGCTTGAATGAGAAATGTGCGCCACCATTTGACGCCAGCGGCTCTG 791
Db 770 CGGGGGAAGATCATGCGCTTGAATGAGAAATGTGCGCCACCATTTGACGCCAGCGGCTCTG 829
QY 792 CCGTGCCACCCCTACGCTGAAGAGCCATTTGAATGTTCACCTTCCAGGCGCGTCCAGGT 851
Db 830 TCAGGCCACCCCTACGCTGAAGAGCCATTTGAATGTTCACCTTCCAGGCGCGTCCAGGT 889
QY 852 CGGCAACCGTGTGCTCAAAAGCCATGTGAACAATGCTTCAAACTAGCATGAGAGT 911
Db 890 GGGGGAACCGTGTGCTCAAAAGCCATGTGAACAATGCTTCAAACTAGCATGAGAGT 949
QY 912 GGGGCTGTGCTGAGAGGCTTATGCGCAGAGAGGCTGAGACCCACGCGGCGCATATCAAG 971
Db 950 GGGTGTGTGTGAGAGGCTTATGCGCAGAGAGGCTGAGACCCACGCGGCGCATATCAAG 1009
QY 972 TGCCTTATATGACTTTGTGTGCTGAGAGCGAGATGACCAAGCCCAAGTGTGCGCGTGTAT 1031
Db 1010 CGCCTTATATGACTTTGTGTGCTGAGAGCGAGATGACCAAGCCCAAGTGTGCGCGTGTAT 1069
QY 1032 TCGGCCCCAGCGCGGATGTGAGCGCGGCTACCGAGAGGCGCAATGCGAGAAAGAT 1091
Db 1070 TCGTCCCCAGCGCGGATGTGAGCGCGGCTACCGAGAGGCGCAATGCGAGAAAGAT 1129
QY 1092 CCGGCTGAGACAGAAATACCTTGTGTCTTGAACAGGCAAAATGTGCGCGTGTGTCC 1189
Db 1130 CCGGCTGAGACAGAAATACCTTGTGTCTTGAACAGGCAAAATGTGCGCGTGTGTCC 1189
QY 1152 CTGGGACCTTAGCAACAGAGTGTACTGAGCTAACATTAACCTTCTCTTCAAGATGCT 1211
Db 1190 CTGGGACCTTAGCAACAGAGTGTACTGAGCTAACATTAACCTTCTCTTCAAGATGCT 1249
QY 1212 TGTGGCAAGACCAATGTGGTGTCTGCTCGAGATCATGCAAGTCCGCTGTACACTCT 1271
Db 1250 CATGGCAAGACCAATGTGGTGTCTGCTCGAGATCATGCAAGTCCGCTGTACACTCT 1309

QY		1272	GGAGGATGACAAAGTTCCTCCTCCCTCCAATTGAATGGTGTGCATGGATGCGAGGCCA	1331
Db		1310	AGAAGAGGAC---TTCTCTCTCTTCACTTGGAATGGTGTAAATGTGATGCCGCCA	1366
QY		1332	GGCCTTCTGCTGCTCTCGCATCGACTCGTCAGAGGCCAGAGTGGACAACGACTACCGAG	1391
Db		1367	GGTCTTTAGCTGCTGTCTACAGACTCTGCCGAGGACCAAGAGTGGGCAACAGACTTAACCGGAG	1426
QY		1392	CGTGGAGCTAGTAGCAGCAGGTGAGCAGAGAACAGAGCCATCTAACACGTCACCACCTCTGC	1451
Db		1427	TGTGGACCTGTGTGAGCAAGTGAATGAGGAAGAAGCCATCTAACACGTCATCACACCCCGC	1486
QY		1452	CCTGGAGAGTGCACCAACAAGCCCCCAGACTTGTGATCCTGGCTCTCGAGCGGAGACCTTG	1511
Db		1487	CCTTAGCGGGGAACACCAAGCCCCCAGACTTGTGATCTGGCTCTTAGCGGAACTTG	1546
QY		1512	TGACCAATGGGAGACCCCTATGTGATCGCGCTGAGGTGGTTCGCTGCCACACACCGAGA	1571
Db		1547	TGAACAATGGGAGACCCCTATGTGATTCGCGCTGAGGTGGTTCGCTGCCACACACATGA	1606
QY		1572	GAGCCCAAGATGACAGACGCGGAGAGACCCCTGCTCAGCGGCTTCGCTCGGCGGAGGG	1631
Db		1607	GACACCCGAAATACCAAGCGTGGGAGACTCTGTTCAGGCTTGTGTGTGTGGCTGAGGG	1666
QY		1632	GGACCAGCTGACCAAG	1647
Db		1667	GGACCAAGTGACTAAG	1682
RESULT 9				
LOCUS		BC042492	5606 bp mRNA linear ROD 07-OCT-2003	
DEFINITION		Mus musculus thioesterase, adipose associated, mRNA (cDNA clone		
VERSION		MG:25974 IMAGE:4240966), complete cds.		
KEYWORDS		BC042492 GI:27503494		
SOURCE		MGCL		
ORGANISM		Mus musculus (house mouse)		
REFERENCE				
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 5606) Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Ditchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stepleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ueding,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramsen,R.D., Mullany,S.U., Bosak,S.A.C., McEwan,P.J., McMernan,K.U., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulvik,S.W., Vallalun,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Faney,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shetchenko,Y., Bouford,G.G., Blakesley,R.W., Touchman,J.W., Green,B.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,V.S., Krzywinski,M.I., Skalka,U., Smalls,D.E., Schereh,A., Schein,J.R., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
JOURNAL				
MEDLINE		22388257		
PUBMED		12477932		
REFERENCE		2 (bases 1 to 5606)		
AUTHORS		Straussberg,R.		
TITLE		Direct Submission		
JOURNAL		Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGCL), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		

REMARK
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowls, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

FEATURES
 source
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 31 Row: f Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 22165403.
 Location/Qualifiers
 1..5606
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="MGC:25974 IMAGE:4240966"
 /tissue_type="Kidney, normal, 5 month old male mouse."
 /clone_lib="MCI CGAP_Kid14"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..5606
 /gene="Thea"
 /note="synonyms: 1110020M10R1k, BFIT, BFITL, MGC25974,
 Thea1"
 /db_xref="locusid:329910"
 /db_xref="MGI:1913736"
 51..1835
 /codon_start=1
 /product="thioesterase, adipose associated"
 /protein_id="AA42492.1"
 /db_xref="GI:27503495"
 /db_xref="locusid:329910"
 /translation="MTQNGCNHLRPGFASMSNRTSRKSIHSDGDPPTMARGEGR
 NPTREVQMSQLVPECHTNHKGELSIGCLIKMIDPTTACLSARHAGCCVYASMDIYED
 HTLSQVQVNLIKAKNKAPEASMEVGQVSEDLCSKQSVKALATPAHREBSKV
 KLAQVPLTEEEKTEHGVAFERRRRMIVADTLDLTHCVIDDLDKDCSNVPAEK
 TRYESVLVLPAPHANQNTFGGQIMAMENAVATIILASRLCHAPTLKALTEMFRRQ
 SOVGDELVLKALVNNAFKHSMEVGVCEAVROAEATORRHINSAFTFVAILDDQCP
 KLPWIRPOEGEGRYREASARKKILDRYLVSCQAEVALSVPMDSNOVLYTARRR
 VSLTKLTMACDNVTLVEISLVRIYLIETPRLSEPHLEMYVNVDAQVFOILSDIRRR
 EMKHKRSVELVQVDEDDAIHYIVISFALSGNTKRPDPVILASRRKCDNGDPYVIAL
 RSYLTPHHETPEYQREGELTCSGFCLMRBGSDQMTKSYNQATPGFLNYVTNVSGLS
 SEFYNTTKACESFILNRNDLAPSLQTL"
 189..620
 /note="COG1607; Region: Acyl-CoA hydrolase [lipid
 metabolism]"
 /db_xref="CDI:COG1607"
 684..1142
 /note="COG1607; Region: Acyl-CoA hydrolase [lipid
 metabolism]"
 /db_xref="CDI:COG1607"
 1203..1709
 /note="START; Region: in StAR and phosphatidylcholine
 transfer protein"
 /db_xref="CDD:smart00234"

ORIGIN
 Query Match 66.2%; Score 1230; DB 10; Length 5606;
 Best Local Similarity 85.8%; Pred. No. 1.8e-229;
 Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;

```

Db      50 AATGATTGAAATGTGGGCAACCACTTGGAGAGGGCTTGCTCTATGTTCTCTATG 109
Qy      78 CACATCCCGGAAATGACCTTAGTTCG-----GGAAACGACAGTGCATGCGACG 131
Db      110 CACATCCCGGAAATGATCTCCATCCGAGTCTGAGAGCCCTCTATGATGCAAGAG 169
Qy      132 CGAGGATACCGGAAACCCGACGAGGTGAGATGAGCCAGTGGTCTCCGCGACAC 191
Db      170 TGAAGATACCGGAAACCCGACGAGGTGAGATGAGCCAGTGGTCTCCGCGACAC 229
Qy      192 CAACCAACGTGTGAGCTGAGCGTGGGAGCTGCTCAAGTGAATTGACACGAGCTT 251
Db      230 CAACCAACGTGTGAGCTGAGCATTTGACAGTTGCTCAAGTGAATTGACACGAGCTT 289
Qy      252 CCGTCCCGGAGAGGAGCGCTGGCTCCCTGTCTGCAAGCTTCCATGATGACATCT 311
Db      290 CCAATCAGGAGAGGAGCGATGTCTGTCTGCTGCTGCAAGCTTCCATGATGACATCT 349
Qy      312 TTTTGGACACACCATTAAGTGTGAGCAAGTGGTGAATATCAAGGCCAAGGTGA 371
Db      350 CTTGACCACTACCATTAAGTGTGAGCCAAAGTGGTGAATATCAAGGCCAAGGTGA 409
Qy      372 CTTCAACTCCAGCATGAGAGTGGGCAATCCAGTGGCTTGAGAGACCTTGCTTG 431
Db      410 CTTCAACTCCAGCATGAGAGTGGGCAATCCAGTGGCTTGAGAGATCTTGCTTG 469
Qy      432 GCAGTGAATGTGTGAGAGGCTTGGCCCACTTGGGCGCCCGAGAGATACCAAGT 491
Db      470 GCAGTGAATGTGTGAGAGGCTTGGCCCACTTGGGCGCCCGAGAGATACCAAGT 529
Qy      492 GAAGCTGAAGCAGATCACGCCGCGACAGAAAGAGAGAGATGAGACAGTGTGGCG 551
Db      530 GAAGCTGAAGCAGATCACGCCGCGACAGAAAGAGAGAGATGAGACAGTGTGGCG 589
Qy      552 TGAAGCGCGCGCAGTCCGCTTGTCTATGACAGACCATCAAGACCTTGCGCA 611
Db      590 TGAAGCGCGCGCAGTCCGCTTGTCTATGACAGACCATCAAGACCTTGCGCA 649
Qy      612 CGCCATTGAGGGCGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 671
Db      650 TGTCACTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 706
Qy      672 CCGTGTGAGAGAGTGTGAGAGTGTCTGCTCCGACGCGCATCAAGAGGCAACCTT 731
Db      707 CCGAGTGAAGAGTGTGAGAGTGTCTGCTCTCTACGCGCATCAAGAGGCAACCTT 766
Qy      732 TGGGGGCCAGATCATGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 791
Db      767 CGGGGGAGAGATCATGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 826
Qy      792 CCGTCCCGACCCCTAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 851
Db      827 TCAGGCCACACCTTACGCTCAAGGCCATGAGAGTGTCTCAATTCGAGGCCGCTCA 886
Qy      852 CGGCGACCGTGTGCTCAAGGCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 911
Db      887 GGGGGACCGTGTGCTCAAGGCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 946
Qy      912 GGGCGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 971
Db      947 GGGGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1006
Qy      972 TGCCCTTATGACCTTGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1031
Db      1007 CGCCTTATGACCTTGTGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066
Qy      1032 TCGGCGCCAGCGCCGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1091
Db      1067 TCGTCCCGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1126
Qy      1092 CCGCTGAGAGAGAGATCATGCTGTCTGTAGAGAGAGAGAGAGAGAGAGAGAGAG 1151

```

```

Db      1127 CCGCCTGAGACAGGAATACCTTGTGTCTGTAGAGAGAGAGAGAGAGAGAGAGAG 1186
Qy      1152 CTGGGACCTTAGCAACAGGTGTACCTGAGACTCAATTAAGTCTCTCTTGAAGTGT 1211
Db      1187 CTGGGACCTTAGCAACAGGTGTACCTGAGACTCAATTAAGTCTCTCTTGAAGTGT 1246
Qy      1212 TGTGGCAGAGACAACTGGGTGTCTCTCGAGAGATCAAGTCAAGTCTCTCTCT 1271
Db      1247 CATTGCCAAGAGCAAACTGGGTGTCTCTCGAGAGATCAAGTCAAGTCTCTCTCT 1306
Qy      1272 GGAGAGTGAACAAGTCTCTCTCTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAG 1331
Db      1307 AGAAGAGAGC---TTCCTCTCTTCACTTGAAGAGTGTGTGAATGTGTGAAGCCG 1363
Qy      1332 GGCCTTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1391
Db      1364 GGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1423
Qy      1392 CGTGGAGCTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1451
Db      1424 TGTGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1483
Qy      1452 CCGTGGAGCTCAACAAGCCCGAGACTTCTGTGATCTCTGAGCTGAGAGAGAGAG 1511
Db      1484 CCGTGGAGCTCAACAAGCCCGAGACTTGTGATCTCTGAGCTGAGAGAGAGAGAG 1543
Qy      1512 TGAAGTGGGAGACCTTATGTATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1571
Db      1544 TGACATGGGAGACCTTATGTATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1603
Qy      1572 GAGCGCAGAGTACAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631
Db      1604 GAGCGCAGAGTACAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1663
Qy      1632 GGACCACTGACCAAG 1647
Db      1664 GGACCACTGACCAAG 1679

```

```

RESULT 10
AK122350
LOCUS Mus musculus mRNA for mKIA0707 protein.
DEFINITION AK122350
ACCESSION AK122350.1 GI:28972356
VERSION FL1_CDNA.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R., Ohara,R., Inamoto,S., Aizawa,H., Yuasa,S.,
Nakajima,D., Nagase,T., Ohara,O. and Koga,H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA
gene: II. The complete nucleotide sequences of 400 mouse
KIAA-homologous cDNAs identified by screening of terminal sequences
of cDNA clones randomly sampled from size-fractionated libraries
JOURNAL DNA Res. 10, 35-48 (2003)
REFERENCE 2
AUTHORS (bases 1 to 5415)
Okazaki,N., Kikuno,R., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(R-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing
FEATURES
SOURCE
1..5415
/organism="Mus musculus"
/mol_type="mRNA"

```

gene
1.5415
/gene="mKIAA0707"
<1.1687
/note="CDS is predicted by in silico analysis. Start codon is not identified."
/codon_start=2
/evidence=not_experimental
/product="mKIAA0707 protein"
/protein_id="BAC65632.1"
/db_xref="GI:28972357"

CDS

/db_xref="taxon:10090"
/clone="mbg07785"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK-"
/gene="mKIAA0707"
<1.1687
/note="CDS is predicted by in silico analysis. Start codon is not identified."
/codon_start=2
/evidence=not_experimental
/product="mKIAA0707 protein"
/protein_id="BAC65632.1"
/db_xref="GI:28972357"
/translation="pPTMARGEGRNPTFEVMSOLVLPCHTHRGELSTGOLLKMTDT
TACLSAERRAGCCPTVASMDDIYFDHITISGVYVNIKAKINRAFNSSMEVGIYVSED
LCEKQMSVCALATFEVARELSKVKLVIPLEBKTHGVAERRRRLVYADIT
KDLITHCVIDDDKDCSNVPAKTEVESEVLELPEPHANQONTFEGQJAMMENVA
TIAASRLCHAHPTLKALIMFHFPGSPQVGRVLKALVNNAPKSMGVCAVRYOE
AETORHINSAFMTEFVLIDKODOPKLPETRPOGEGERYREASARKJRLRKLYLV
SCQAEVYALSVPMDSNOYVLSYVNSLITLMAKQWVLSVERLYLITLEDLS
PHLEMYNVDAOVFOLSLDRRPEMDKHYRSEVLYQVDEDDALYHYISPLASNT
KPDVFLASRRKPCDNDPVIARSLVTLPHETVLPORGETLGFCLMBEDQM
TKVSYNQATPGFLNVTYTVNGSLSEFNTFACESFLDNRNDLAPSLDTL"

ORIGIN

Query Match 63.7% Score 1182.2; DB 10; Length 5415;
Best Local Similarity 86.7%; Pred. No. 3.7e-220;
Matches 1327; Conservative 0; Mismatches 198; Indels 6; Gaps 2;

117 TGCCATGCGACAGCGGAGGATACCGGAAACCCACGAGGTGCAGATGAGCCAGCTGT 176
177 GCTGCCCTGCCACCAACCAACGCTGAGCTGAGCGTGGGACGCTCTCAAGTGAAT 236
67 ACTGCCCTGCCACCAACCAACGCTGAGGAGCTGAGCATTGGAAGTTGCTCAAGTGAAT 126
237 TGACACCAACGAGCTGCTGCCGCGGAGAGGACGCTGCTGCCCTGTGCAAGCTTC 296
127 CGACACCAACGAGCTGCTGCCGCGGAGAGGACGCTGCTGCCCTGTGCAAGCTTC 186
297 CATGATGACATCTATTGTCAGACACCATTAAGTGTGGAACAAGTGTGAATCAAGGC 356
187 TATGATGACATCTATTGTCAGACACCATTAAGTGTGGAACAAGTGTGAATCAAGGC 246
357 CAAGTGAACCGGAGCTTCAACCTCAGCATGAGGTGGGCATCAAGTGTGGCTGAGAGA 416
247 CAAGTGAACCGGAGCTTCAACCTCAGCATGAGGTGGGCATCAAGTGTGGCTGAGAGA 306
417 CTTGCTCTGAGAAAGAGTGAATGTGTGCAAGGCTTGGCCACTTTCGTGGCCACCG 476
307 TCTGTGCTGAGAAAGAGTGAATGTGTGCAAGGCTTGGCCACTTTCGTGGCCACCG 366
477 AGAGATGACCAAGTGAAGTGAAGAGATCAAGCCGCGGACCAAGAGAGAGATGAGA 536
367 GGAAGCTCTCAAGTGAAGTGAAGAGATCAAGCCATGACCAAGAGAGAGATGAGA 426
537 GCACAGTGTGGCGGCTGAGCGCGCGCATGCGCTTGTCTATGCAACACCATCAAGA 596
427 ACATGGGAGTGGCGGCTGAGCGCGCGCATGCGCTTGTCTATGCAACACCATCAAGA 486
597 CTTCTGAGCCACCTGGCGCATTCAGGGGAGTCTGAGAGAGAGAGATGAGCCAGTGT 656
487 TCTCTAAACCACTGTGTCATCCAGAGAGATTTGGA--CAAGACTGCAACAATATAGT 543
657 GCCGCTGAGAGAGCCGCTGTGAGAGTGTGAGACTGTCTGCTGCCACGCAATCA 716
544 GCCAGCGGAGAGAGCCGAGTGTGAGAGTGTGAGACTGTGCTGCTGCCACGCAATCA 603
717 CGAGGCAACACTTTTGGGGGCGCAGATCATGCGCTGATGAGAGATGTGGCCACCATTCG 776

604 TCAGGCGCAATACCTTCGGGGGACAGATCATGCTTGATGAGAAATGTGCCACCATTCG 663
777 AGCCAGCGGAGCTGAGCGTGGCCACCCCTTACGCTGAAGAGCCATTGAATGTTCACCTTCG 836
664 AGCCAGCGGAGCTGTCATCGCCACCCCTTACGCTGAAGAGCCATTGAATGTTCACCTTCG 723
837 AGGCGCGTCCAGAGTGGCGACCGCTGTGCTGCTCAAGCCATCGTGAACAATGCTTCGA 896
724 AGGCGCGTCCAGAGTGGCGAGCGCTGTGCTGCTCAAGCCATCGTGAACAATGCTTCGA 783
897 ACATAGTGTGAGTGGGCTGTGCTGAGAGGCTTATCGCTCAAGAGCTGAGACCCACCG 956
784 GCACAGATGAGAGTGGGCTGTGCTGAGAGGCTTACCGCGCAAGACTGAGACCCACCG 843
957 GCGCCATCATCAACAGTGCCTTTATGACCTTTGATGCTGCTGAGACGAGATGACCGACCA 1016
844 CCGGACATCATCAACAGGCTTTCATGACCTTGTGCTGCTGAGCAAAATGATGACGCTTCA 903
1017 GTTGTGCTGCTGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1076
904 GAAGCTGCTGATGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 963
1077 TGCCAGAAAGAAATGCTGCTGAGACAGAGTACATGCTGCTGCTGAAGAGAGAGT 1136
964 TGCCAGAAAGAAATGCTGCTGAGACAGAGTACATGCTGCTGCTGAAGAGAGAGT 1023
1137 GCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1196
1024 GCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1083
1197 CTCTTGAAGATGCTTGTGAGCAAGCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1256
1084 CTCTTGAAGATGCTTGTGAGCAAGCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
1257 CCGCTGTACATCTGTGAGAGTACAAAGTTCCTCTCTTCCATGAGATGATGTGTGCA 1316
1144 CCGCTGTACATCTGTGAGAGTACAAAGTTCCTCTCTTCCATGAGATGATGTGTGCA 1200
1317 TGTGATGAGAGCGGAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1376
1201 TGTGATGAGAGCGGAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
1377 CAAGCACTTACCGAGGCTGAGAGTACAGAGTACAGAGTACAGAGTACAGAGTACAGAGT 1436
1261 CAAGCACTTACCGAGGCTGAGAGTACAGAGTACAGAGTACAGAGTACAGAGTACAGAGT 1320
1437 CGTACCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1496
1321 CGTACCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1497 GAGCGGAGAGCTTGTGACAAATGAGGAGCCCTATGTCATCGCTGAGAGTGTGAGCT 1556
1381 TAGCGAGAGAGCTTGTGACAAATGAGGAGCCCTATGTCATCGCTGAGAGTGTGAGCT 1440
1557 GCCCAACACCGAGAGAGCGCCAGATGACAGCGGAGAGAGAGCTTGTGCTCAAGCTTCTG 1616
1441 GCCCAACACCGAGAGAGCGCCAGATGACAGCGGAGAGAGAGCTTGTGCTCAAGCTTCTG 1500
1617 CTTGTGCGGAG 1647
1501 TCTGTGAGGAG 1531

RESULT 11.

BC001517 2455 bp mRNA linear PRI 16-SEP-2003
LOCUS Homo sapiens thioesterase, adipose associated, mRNA (cDNA clone
DEFINITION IMAGE:2966784), complete cds.
ACCESSION BC001517
VERSION BC001517.2 GI:33876308
KEYWORDS Homo sapiens (human)
SOURCE

RESULT 12
 AX870465 751 bp DNA linear PAT 17-DEC-2003
 LOCUS AX870465
 DEFINITION Sequence 5370 from Patent EP1074617.
 ACCESSION AX870465
 VERSION AX870465.1 GI:40025328
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 Primers for synthesizing full-length cDNA and their use
 Patent: EP 1074617-A 5370 07-FEB-2001;
 Research Association for Biotechnology (ur)

FEATURES
 source
 1. .751
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 34.3%; Score 636.4; DB 6; Length 751;
 Best Local Similarity 97.8%; Pred. No. 9.6e-114;
 Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTCTGTCTTCCAAACCGC 78
 DB 83 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTCTGTCTTCCAAACCGC 142

QY 79 ACATCCCGGAAGTACAGCTTACGTCGCGGGGAACGACAGTGCACATGAGAGAGCGGAGGGA 138
 DB 143 ACATCCCGGAAGTACAGCTTACGTCGCGGGGAACGACAGTGCACATGAGAGAGCGGAGGGA 202

QY 139 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACACCAACCA 198
 DB 203 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACACCAACCA 262

QY 199 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAAGTGAACACGAGCTTGCTGCTGCTC 258
 DB 263 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAAGTGAACACGAGCTTGCTGCTGCTC 322

QY 259 GCGGAGGAGCAGCGTGCCTGCCCTGTGTCAACAGCTTCCATGATGATCATATTTAG 318
 DB 323 GCGGAGGAGCAGCGTGCCTGCCCTGTGTCAACAGCTTCCATGATGATCATATTTAG 382

QY 319 CACACCAATTAGTTGGACAAGTGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAAC 378
 DB 383 CACACCAATTAGTTGGACAAGTGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAAC 442

QY 379 TCCAGCATGAGAGTGGGATTCAGGTGCTCGAGGACCTGTGCTTGAAGCACTG 438
 DB 443 TCCAGCATGAGAGTGGGATTCAGGTGCTCGAGGACCTGTGCTTGAAGCACTG 502

QY 439 AATGTGTCAAGGCTTGGCCACCTTGTGCTGCCCGCGAGATCAACCAAGTGAAGCTG 498
 DB 503 AATGTGTCAAGGCTTGGCCACCTTGTGCTGCCCGCGAGATCAACCAAGTGAAGCTG 562

QY 499 AAGCAGATCAAGCGCGGACAGAGAGAGAGATGAGCAGTGTGTGGCGGTGAGGCG 558
 DB 563 AAGCAGATCAAGCGCGGACAGAGAGAGAGATGAGCAGTGTGTGGCGGTGAGGCG 622

QY 559 CGGGCATGCGCTTGTCTATGACACCATCAAGGACCTTCTGGCCACTGCGCAT 618
 DB 623 CGGGCATGCGCTTGTCTATGACACCATCAAGGACCTTCTGGCCACTGCGCAT 682

QY 619 CAGGCGCATCTGG-AGAGCAGAGCTGAGCGCGCATGTTGCCGCTGAGAGACCCGT 677
 DB 683 CAGGCGCATCTGAGAGAGAGAGAGAGCTGATCCGATGTTGCCGCTTANAAAAACCCGT 742

QY 678 CGAGAGTGT 686
 DB 743 NGAAAAATGT 751

RESULT 13
 BD150527 751 bp DNA linear PAT 17-JAN-2003
 LOCUS BD150527
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD150527
 VERSION BD150527.1 GI:27856285
 KEYWORDS JP 2002191363-A/5370.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 751)
 Oca,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
 Primers for synthesizing full-length cDNA and use thereof
 Patent: JP 2002191363-A 5370 09-JUN-2002;
 HELIX RESEARCH INSTITUTE

FEATURES
 source
 1. .751
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 34.3%; Score 636.4; DB 6; Length 751;
 Best Local Similarity 97.8%; Pred. No. 9.6e-114;
 Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTCTGTCTTCCAAACCGC 78
 DB 83 ATGATCCAGAAATGTCGGAATCACTGCGACGCGGGCTTGCCCTGTCTGTCTTCCAAACCGC 142

QY 79 ACATCCCGGAAGTACAGCTTACGTCGCGGGGAACGACAGTGCACATGAGAGAGCGGAGGGA 138
 DB 143 ACATCCCGGAAGTACAGCTTACGTCGCGGGGAACGACAGTGCACATGAGAGAGCGGAGGGA 202

QY 139 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACACCAACCA 198
 DB 203 TACCGGAACCCCAACGAGGTGACAGTATGAGCAGCTGTGCTGCCCTCCACACCAACCA 262

QY 199 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAAGTGAACACGAGCTTGCTGCTGCTC 258
 DB 263 CGTGTGAGCTGAGCGTTCGCGGACGCTGCTCAAGTGAACACGAGCTTGCTGCTGCTC 322

QY 259 GCGGAGGAGCAGCGTGCCTGCCCTGTGTCAACAGCTTCCATGATGATCATATTTAG 318
 DB 323 GCGGAGGAGCAGCGTGCCTGCCCTGTGTCAACAGCTTCCATGATGATCATATTTAG 382

QY 319 CACACCAATTAGTTGGACAAGTGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAAC 378
 DB 383 CACACCAATTAGTTGGACAAGTGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAAC 442

QY 379 TCACACATGAGAGTGGGATCCAGGTGGCTCGAGAGACCTGTGCTTGAGAGACAGTGG 438
 Db 443 TCACACATGAGAGTGGGATCCAGGTGGCTCGAGAGACCTGTGCTTGAGAGAGAGTGG 502
 QY 439 AATGTGTGACAGGCTTGGCCACCTTGTGGCCCGCCGAGAGATCAACCAAGGTGAAGCTG 498
 Db 503 AATGTGTGACAGGCTTGGCCACCTTGTGGCCCGCCGAGAGATCAACCAAGGTGAAGCTG 562
 QY 499 AAGCAGATCAGCCCGCCGAGAGAGAGAGATGAGACCAAGTGTGGCGGCTGAGCCG 558
 Db 553 AAGCAGATCAGCCCGCCGAGAGAGAGAGAGATGAGACCAAGTGTGGCGGCTGAGCCG 622
 QY 559 CGGCGCATGCGCCTTGTGTATGACACACATCAAGACCTCTGAGCCCACTGGCCGAT 618
 Db 623 CGGCGCATGCGCCTTGTGTATGACACACATCAAGACCTCTGAGCCCACTGGCCGAT 682
 QY 619 CAGGCGCATGTGG-AGAGCAGAGACTGTAGCCGCGATGTGGCGGCTGAGAGACCGGCTG 677
 Db 683 CAGGCGCATGTGGAG 742
 QY 678 GGAGAGTGT 686
 Db 743 NGAAATGT 751
 RESULT 14
 AB078618 1689 bp mRNA linear ROD 24-JAN-2002
 LOCUS Mus musculus mCACH-1 mRNA for cytosolic acetyl-CoA hydrolase,
 DEFINITION complete cds.
 ACCESSION AB078618
 VERSION AB078618.1 GI:18307691
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS Suematsu, N., Okamoto, K. and Isohashi, F.
 TITLE Cytosolic acetyl-CoA hydrolase
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 1689)
 AUTHORS Suematsu, N., Okamoto, K. and Isohashi, F.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-2002) Naoya Suematsu, St. Marianna University
 School of Medicine, Department of Biochemistry, Miyamae-ku, Sugo
 2-16-1, Kawasaki, Kanagawa 216-8511, Japan
 (E-mail: n2sue@marianna-u.ac.jp, Tel:81-44-977-8111 (ex.3525),
 Fax:81-44-976-7553)
 FEATURES
 source
 1..1689
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /sex="male"
 /tissue_type="liver"
 1..1689
 /gene="mCACH-1"
 8..1678
 /gene="mCACH-1"
 /codon_start=1
 /product="cytosolic acetyl-CoA hydrolase"
 /protein_id="BAB84021.1"
 /db_xref="GI:18307692"
 /translation="MESVAPGEVLMGQATIQPAHADRGLSAGQLKMDTDLAA
 EKHAGISCVTASMDIIFEDFARIQGITIRAKVTAFASMEISIKYIVQDKTGLG
 KLICVASTFPAKGVKKEKYLKPVLOTBOVEHNLASRRKRLQHEHNTPNNIK
 ESSRFSDSICNBEETATMTGTSYOSIEIVLPHANHGNTFGQIIMAMETVATTA
 SRUCHGHPFLKSVMPKFRGSRSTYGDRLVRSATYNNNTFONSVEGVVEADCECMAB
 GGGHINSAFLINAVDDQELTFPRIQPSKDFRYYGALARRIRLGRKXVISH
 KKEVPLSAQWDISKKGLSNTNVALKNLASKSGMEITTLLEKIKYITLEQDAISVK

ORIGIN
 Query Match 23.4%; Score 434.8; DB 10; Length 1689;
 Best Local Similarity 57.9%; Pred. No. 1.6e-74;
 Matches 854; Conservative 0; Mismatches 607; Indels 15; Gaps 4;
 VETKNGSPAHIAVHLISDLTRPLMDPHYISGEVLDVSEDDOITYITCSYVNGDKPK
 DEIVYVSRKRLKDNNTTVAIRSVYLVSPVSPQIINSEVLCAGFLIOAVDSNCTV
 TYLNQMSDILPYFAGNIGWSKSIIEAASCIKFIENATPGILKSLV"
 154 GAGGTGACAGATGAGCCAGTGTGTGCTCCCTCCACACCAACCAAGTGTGAGCTGAGC 213
 Db 32 GAGGTGCTCATGAGCCAGGCCATCCAGCGGCTCATGCGCATCCCGCGGCGAGCTGAGC 91
 QY 214 GTGCGGCGAGCTGCTCAAGTGAATTGACACCAAGCGTCTGCTCCGCGGAGAGAGACGCT 273
 Db 92 GAGGCGAGCTGCTCAAGTGAATTGACACCAAGCGTCTGCTCCGCGGCGGAGAGAGAGCT 151
 QY 274 GAGTCCCGCTGTGTGACAGCTTCCATGATGATCATCTTATTTGAGGACACCAATTAGTGT 333
 Db 152 GGGATTTCTGTGTGACAGCCCTCCATGATGATCATCTTGTGAGGACACAGCGAATT 211
 QY 334 GGAAGAATGATGATATTAAGGCCAAGGTGAACCGGGCTTGAACCTCACTGAGATGAGGTG 393
 Db 212 GGACAAATTAATTAACCATGAGAGCAAAAGTGAAGTGAAGCGCTTCAAGACATGAGATC 271
 QY 394 GGCATCCAGGTGCGCTCGGAGACCTGTGCTGTGAGAAAGAGTGAATGTGCAAGGCC 453
 Db 272 AGATCAAGGTGATATGATGTCAGAGCAAGTGTACAGGCAATCCAGAACTCTGTGCGTGT 331
 QY 454 TTGGCCACCTTGTGTGCGCCCGCCGAGAGATCA---CCAAGTGAACCTGAGAGAGATCAAG 510
 Db 332 TTCTCTACGTTTGTGATGCTTAACCAAGTGTGCAAAAGAAAGTGTCACTTAAACCTGTCTG 391
 QY 511 CCGCGGACAGAAAGAGAGAGATGAGCAAGTGTGCGGCTGAGAGCGCGGCGATGCGC 570
 Db 392 CTTCAACAGAGAGCAAGAAACAAGTGAAGCAATCTGCTTCAAGAGAGAAAGTCCGA 451
 QY 571 CTTGTCTATGAGACACCATGAGAGACCTCTGTGCGCACTGCGCCATTCAGGCGATCTG 630
 Db 452 CTGAGCATGAGAACACCTTCAACCAATTAATGAAGAGAGAGAGATTCAGCATTC 511
 QY 631 GAGAGCAGAGACTGTAGCCGATGTGCGGCTGAGAGACCCGCTGAGAGAGTGTGAG 690
 Db 512 ATTTGTATGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 571
 QY 691 CTGTCTCTGCTCCCGACGCAATGACAGAGGCAACACTTTGAGGCGCAATGAGCC 750
 Db 572 CTGTCTCTGCTCCCGACGCAATGACAGAGGCAACACTTTGAGGCGCAATGAGCC 631
 QY 751 TGGATGAGAGATGTGGCCAGCATTTGAGAGCGGCGCTCTGCGGCCACCGCTTACGCTG 810
 Db 632 TGGATGAGAGAGATGTGGCCAGCATTTTCTGAGAGCGGCGCTCTGCGGCCACCGCTTACG 691
 QY 811 AAGGCCATTTGAATATTTCCACTTCCAGAGCCCGCTCCAGAGTGTGAGAGAGTGTGCTC 870
 Db 692 AAGTCTGTGATATGTTAAATTTCCGGGAGAGCTGCAAGTGTGAGAGAGAGCTTCTTC 751
 QY 871 AAAGCCATCGTGAACATGCTTCAAAATATGATGAGAGTGTGAGAGAGAGAGAGAGAG 929
 Db 752 AGTCCATAGTCAACCAACCTTTCAAGAACAGTGTGAGAGTGTGAGAGAGAGAGAGAG 811
 QY 930 -----CTATGCCAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 984
 Db 812 TTTAATCTGTGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
 QY 985 TTTGTGTCTGTGAGAGAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1044
 Db 872 TACATATGCTGTGATGACAG 931
 QY 1045 GGCATGTGTAGAGCGGCTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104
 Db 932 AAGACGATTTTTCCTGTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991

QY	865	GTGCTCAAGCCATGCTGTAACAATCCCTTCAAAAGTAACTGAGAGTGGGCGTGCGCTG	924
Db	759	GCTTCAATGCCATGCTGTCAACAACCTTTTCAAAACAGTGTGGAAGTGGCGTACGTGT	818
QY	925	GAAGC-----CTATGCCAGGAGGCTGAGACCACCGGCGCCATCATCAACAGTSCCTT	978
Db	819	GAGGCGCTTGAAGTGTGGGAGTGGGCGGAGCCGAAGGCCGCAATCAACAGCGCTTTT	878
QY	979	ATGACCTTTGTGTCTCTGGAGCGAGATGACCAAGCCCCAGTTGTCTGCTCGATTCGAGCC	1038
Db	879	CTCATTTTCAATGCTGTGTATGACACAGAGGAACCTCATACCTTTCCAGAAATCCAGCCC	938
QY	1039	CAGCCCCGACATGCTGAGCGGCGGTACCGAGAGCGCAGTGGCAGAAACAGTCCGCTG	1098
Db	939	ATTTCAAAGGATGATTTTCGACGGTACCAAGGAGCTATTGCAAGGAGAAATTCGCTG	998
QY	1099	GACAGGAAGTACATCGTCTCTGTAAAGCAGAGAGTGCCCTCTCCGTCGCCCTGGAGC	1158
Db	999	GGCAGGAATATGTATTTTC---CACAAAGAAAGTCCCACTTTGGCAGCAACATGGAT	1055
QY	1159	CCTAGCAACCAAGGTGTACCTGACTCAATACGTCTCTCCCTTGAAGATCTTGTGGCC	1218
Db	1056	ATAAGCAAAAAAGGATCCATTAAGTAAACCAATGTGAAGCTCTCAAAAAATCTGGCATCC	1115
QY	1219	AAGGCAACCTGGGTGCTGCTCTCGAGATTCAGTCAAGTCCGCGCTGTACCTCTGAGGAT	1278
Db	1116	AAAAAGCGTTGGAGATACCAACCACTTGAAGAAATPAAAAATATATACACTGGAAGG	1175
QY	1279	GACAAGTTCCTCTCTTCCACATGAGATGTGTGTCATGTGATGTGACGCCCAAGCCCTTC	1338
Db	1176	CAGATGTCATATCTGTAAAGGTGGAAAAACAGGTGTGAATCCAGCCCGGTGGCTTAT	1235
QY	1339	CTGCTGCTCTTGGACCTGCGTCAAGAGCCAGAGTGGGACAAGCACTACCGAGCGTGGAG	1398
Db	1236	CATCTCTGTGCGACTTCACAAAAGCGCCCTTTATGGGACCCCCATTCATATCTTTGAA	1295
QY	1399	CTAGTGCAGCAGGTAAAGAGAGAGAGGCCATCTACACAGTCCACAGCGCTGCCCTCGGA	1458
Db	1296	GTCATATGACCAAGGTAGTGAAGAGCATCAGATATATTAACATCACTGCTGTGGTGAAT	1355
QY	1459	GGTCACACAAAGCCCCAGAGCTTCGTGATCTCTGCGCTCGAAGGCGGAAGCCTGTGACAT	1518
Db	1356	GGAGACAAA---CCCAAGGACTTGTGTATCTTGTGTCACAAAGAAAGGCCCTCAAGAC	1412
QY	1519	GGGAGCCCTTATGTCAATCGCGCTGAGGTGCTCAAGCTGCCACACACGAGAGACGCCA	1578
Db	1413	GACAAACCTAACATCGTGGCACATATGTCAGTGTACTGCGCGTGTGCCCTCATCTCCA	1472
QY	1579	GAGTACAGAGCGCGGAGAGACCTCTGCTCAAGGCTTCTGCGCTCTGGC	1624
Db	1473	CAGTACATCAAGACCAAGGTATTTGTGTGGGTCTCTCATTCACGC	1518

> O < IntellGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-10-055-624b-2.res made by tport on Tue Aug 104 10:14:57-PST.

Query sequence being compared: US-10-055-624B-2 (1-607)
Number of sequences searched: 3
Number of scores above cutoff: 3

Results of the initial comparison of US-10-055-624B-2 (1-607) with:
File : US10055624B.pep

```

100-
N -
U -50-
M -
B -
E -
R -
O -
F -10-
S -
E -5-
Q -
U -
N -
M -
C -
E -
S 0-----
SCORE 0 67 135 202 270 337 405 472 540 607
STDEV -2 -1 -1 * * *

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 500
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 496 Median 327 Standard Deviation 149.90
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 1795
Number of sequences searched: 3
Number of scores above cutoff: 3

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Init. Opt.	Length	Score	Score	Sig.	Frame
---------------	-------------	------------	--------	-------	-------	------	-------

The list of other best scores is:

1. US-10-055-624B-2 Sequence 2, Application U 607 607 607 0.74 0

Sequence Name	Description	Length	Score	Score	Sig.	Frame
---------------	-------------	--------	-------	-------	------	-------

2. US-10-055-624B-4 Sequence 4, Application U 594 557 563 0.41 0
**** 1 standard deviation below mean ****

3. US-10-055-624B-6 Sequence 6, Application U 594 326 529 -1.13 0

1. US-10-055-624B-2 (1-607)
US-10-055-624B-2 Sequence 2, Application US/10055624B

Initial Score = 607 Optimized Score = 607 Significance = 0.74
Residue Identity = 100% Matches = 607 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X 10 20 30 40 50 60 70
MIQVGNHILRGILASVFSNRTSRKSLRAGNDSAMADGEGYNPTFVQMSQVLPCHTNQRGELISVGLIKM
X 10 20 30 40 50 60 70
X 10 20 30 40 50 60 70
X 10 20 30 40 50 60 70

```

```

80 90 100 110 120 130 140
IDTTACTSABRNAGCCPVTASMDIYFEHTISVGVYNIKAKVNRKFNSSMEVGIQVASEDLCSKQWNVCK
IDTTACTSABRNAGCCPVTASMDIYFEHTISVGVYNIKAKVNRKFNSSMEVGIQVASEDLCSKQWNVCK
80 90 100 110 120 130 140

```

```

150 160 170 180 190 200 210
ALATFVARREITVKKQITPRTEEEKMEHSVAERRRMELVYADITKDLANCAIQGLDESRCRMPAE
ALATFVARREITVKKQITPRTEEEKMEHSVAERRRMELVYADITKDLANCAIQGLDESRCRMPAE
150 160 170 180 190 200 210

```

```

220 230 240 250 260 270 280
KTRVESVELVLPNHNQNTFGQIMAMMENVAITIASRLCAHPTLKAIEHFHRRGSPQVGRVYLKAIY
KTRVESVELVLPNHNQNTFGQIMAMMENVAITIASRLCAHPTLKAIEHFHRRGSPQVGRVYLKAIY
220 230 240 250 260 270 280

```

```

290 300 310 320 330 340 350 360
NNAFKISMEEVGCVEAYROEAEHRRHINSAPMTFVLDADDDQQLIPWIRPOGDERRRYREASARKIRL
NNAFKISMEEVGCVEAYROEAEHRRHINSAPMTFVLDADDDQQLIPWIRPOGDERRRYREASARKIRL
290 300 310 320 330 340 350 360

```

```

370 380 390 400 410 420 430
DRKYIVCKQTEVPLSVWDPNSQVYLSYNNVSSLKMLVAKDNVLSSEISQVRLYTLDDKFLSPFMENVY
DRKYIVCKQTEVPLSVWDPNSQVYLSYNNVSSLKMLVAKDNVLSSEISQVRLYTLDDKFLSPFMENVY
370 380 390 400 410 420 430

```

```

440 450 460 470 480 490 500
HVDAAQAFLLISDLRORPEWDKHYRSVELVQVDEDDAIYHTVSPALGHTKPPDFTILASRRKPCDNDPY
HVDAAQAFLLISDLRORPEWDKHYRSVELVQVDEDDAIYHTVSPALGHTKPPDFTILASRRKPCDNDPY
440 450 460 470 480 490 500

```

```

510 520 530 540 550 560 570
VIALRSVTLPTFRRETPYRGETLCSGFCIMRREDQTLKCCWVNRSLTELVSAGSFGYWGLESRRGRSDG
VIALRSVTLPTFRRETPYRGETLCSGFCIMRREDQTLKCCWVNRSLTELVSAGSFGYWGLESRRGRSDG
510 520 530 540 550 560 570

```

```

580 590 600
WNGKLAGHLSLTKAIPVAKINSRGLQDT X
WNGKLAGHLSLTKAIPVAKINSRGLQDT X
580 590 600

```

2. US-10-055-624B-2 (1-607)

US-10-055-624B-4 Sequence 4, Application US/10055624B

Initial score	=	557	Optimized score	=	563	Significance	=	0.41
Residue Identity	=	92%	Matches	=	547	Mismatches	=	46
Gaps	=	0	Conservative Substitutions	=			=	1

X	10	20	30	40	50	60	70																																																						
M	Q	N	G	N	H	R	R	G	L	A	S	V	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	G	R	N	T	E	V	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K		
X	10	20	30	40	50	60	70																																																						
M	Q	N	G	N	H	R	R	G	L	A	S	V	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	G	R	N	T	E	V	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K		
X	10	20	30	40	50	60	70																																																						
M	Q	N	G	N	H	R	R	G	L	A	S	V	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	G	R	N	T	E	V	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K		
X	10	20	30	40	50	60	70																																																						
M	Q	N	G	N	H	R	R	G	L	A	S	V	F	S	N	R	S	R	K	S	A	L	A	R	G	N	D	S	A	M	A	D	G	E	G	R	N	T	E	V	O	S	V	L	P	C	H	T	N	O	R	G	L	S	T	G	L	L	K		
80	90	100	110	120	130	140																																																							
I	D	T	T	A	C	L	S	E	R	H	A	G	C	P	C	T	A	S	M	D	I	F	E	H	T	I	S	V	G	O	V	N	I	A	K	N	R	A	F	N	S	M	E	V	I	O	A	S	E	D	I	C	S	E	K	O	N	M	V	C	K
80	90	100	110	120	130	140																																																							
I	D	T	T	A	C	L	S	E	R	H	A	G	C	P	C	T	A	S	M	D	I	F	E	H	T	I	S	V	G	O	V	N	I	A	K	N	R	A	F	N	S	M	E	V	I	O	A	S	E	D	I	C	S	E	K	O	N	M	V	C	K
150	160	170	180	190	200	210																																																							
A	L	A	T	V	A	R	R	E	I	T	K	V	L	K	O	I	T	R	T	E	E	K	H	S	V	A	E	R	R	M	L	V	A	D	I	T	K	O	L	L	A	N	C	A	I	O	G	D	L	E	S	R	D	S	R	M	V	A	E		
150	160	170	180	190	200	210																																																							
A	L	A	T	V	A	R	R	E	I	T	K	V	L	K	O	I	T	R	T	E	E	K	H	S	V	A	E	R	R	M	L	V	A	D	I	T	K	O	L	L	A	N	C	A	I	O	G	D	L	E	S	R	D	S	R	M	V	A	E		
150	160	170	180	190	200	210																																																							
A	L	A	T	V	A	R	R	E	I	T	K	V	L	K	O	I	T	R	T	E	E	K	H	S	V	A	E	R	R	M	L	V	A	D	I	T	K	O	L	L	A	N	C	A	I	O	G	D	L	E	S	R	D	S	R	M	V	A	E		
220	230	240	250	260	270	280																																																							
K	T	R	V	S	S	V	L	P	P	H	A	N	O	G	N	P	F	G	O	I	M	A	M	E	N	V	A	T	I	A	S	L	C	A	H	P	T	L	K	A	I	M	F	H	R	G	S	O	G	D	L	V	L	K	A	I	V				
220	230	240	250	260	270	280																																																							
K	T	R	V	S	S	V	L	P	P	H	A	N	O	G	N	P	F	G	O	I	M	A	M	E	N	V	A	T	I	A	S	L	C	A	H	P	T	L	K	A	I	M	F	H	R	G	S	O	G	D	L	V	L	K	A	I	V				
230	240	2																																																											

[illegible]

> O <
O| 0 IntellGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-10-055-624b-4.res made by tport on Tue Aug 104 10:16:24 PST.

Query sequence being compared: US-10-055-624B-4 (1-594)
Number of sequences searched: 3
Number of scores above cutoff: 3

Results of the initial comparison of US-10-055-624B-4 (1-594) with:
File : US10055624B.pep

```

100-
-
N
-
U 50-
-
M
-
B
-
E
-
R
-
O
-
F 10-
-
S
-
B 5-
-
O
-
U
-
N
-
E
-
C
-
S
-
S
-
0
-
SCORE 0 66 132 198 264 330 396 462 528 594
STDEV -2 -1 1

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 500
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean Median Standard Deviation
491 323 147.97
Times: CPU
00:00:00.00 Total Elapsed
00:00:00.00

Number of residues: 1795
Number of sequences searched: 3
Number of scores above cutoff: 3

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

Sequence Name Description Init. Opt. Length Score Score Sig. Frame

The list of other best scores is:

1. US-10-055-624B-4 Sequence 4, Application U 594 594 594 0.70 0

Sequence Name Description Length Score Score Init. Opt. Sig. Frame

2. US-10-055-624B-2 Sequence 2, Application U 607 559 565 0.46 0

3. US-10-055-624B-6 Sequence 6, Application U 594 322 558 -1.14 0

1. US-10-055-624B-4 (1-594)
US-10-055-624B-4 Sequence 4, Application US/10055624B

Initial Score = 594 Optimized Score = 594 Significance = 0.70
Residue Identity = 100% Matches = 594 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X 10 20 30 40 50 60 70
MIONVGNHRLRGGLASVFSNRTSRKSLRAGNDSAMADGEGYRNPTFVMSQVLVPCHTNORGELSVGLKMW
MIONVGNHRLRGGLASVFSNRTSRKSLRAGNDSAMADGEGYRNPTFVMSQVLVPCHTNORGELSVGLKMW
X 10 20 30 40 50 60 70

80 90 100 110 120 130 140
IDTTACLSAERHAGCCVTAASMDIYFEHTTISVGQVYNIKAKYNRAFNSSMEVGIQVASEDLCSKQWVCK
IDTTACLSAERHAGCCVTAASMDIYFEHTTISVGQVYNIKAKYNRAFNSSMEVGIQVASEDLCSKQWVCK
IDTTACLSAERHAGCCVTAASMDIYFEHTTISVGQVYNIKAKYNRAFNSSMEVGIQVASEDLCSKQWVCK
80 90 100 110 120 130 140

150 160 170 180 190 200 210
ALATFVARRRITVKLKOITPRTEEEKMEHSVAERRRMVYADITKOLLANCAIQGLDLSRDSCMWPAE
ALATFVARRRITVKLKOITPRTEEEKMEHSVAERRRMVYADITKOLLANCAIQGLDLSRDSCMWPAE
ALATFVARRRITVKLKOITPRTEEEKMEHSVAERRRMVYADITKOLLANCAIQGLDLSRDSCMWPAE
150 160 170 180 190 200 210

220 230 240 250 260 270 280
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTAAASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTAAASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV
KTRVESVEVLPPHANHQNTFGQIMAMMENVATTAAASRLCAHPTLKAIEHFHFRGSPQVGRVLVKAIV
220 230 240 250 260 270 280

290 300 310 320 330 340 350 360
NNAFKSMEEVGVCEAARQAEHRHINSAPMTFVVLADDDQPLLPMTRPQDGERERYREASARKIRL
NNAFKSMEEVGVCEAARQAEHRHINSAPMTFVVLADDDQPLLPMTRPQDGERERYREASARKIRL
NNAFKSMEEVGVCEAARQAEHRHINSAPMTFVVLADDDQPLLPMTRPQDGERERYREASARKIRL
290 300 310 320 330 340 350 360

370 380 390 400 410 420 430
DRKTVSCKQTEVEPLSVMPDPSNQVYLSYNNVSLKMLVAKDMWVLSSEISOVRLYTLLEDDKFLSFHMEVY
DRKTVSCKQTEVEPLSVMPDPSNQVYLSYNNVSLKMLVAKDMWVLSSEISOVRLYTLLEDDKFLSFHMEVY
DRKTVSCKQTEVEPLSVMPDPSNQVYLSYNNVSLKMLVAKDMWVLSSEISOVRLYTLLEDDKFLSFHMEVY
370 380 390 400 410 420 430

440 450 460 470 480 490 500
HVDAAQAFLLSLDLRQRPMDKHYRSEVELVQVDEDDAIHVTSFPAIGGHTKPODFVILASRRKPCNGDPY
HVDAAQAFLLSLDLRQRPMDKHYRSEVELVQVDEDDAIHVTSFPAIGGHTKPODFVILASRRKPCNGDPY
HVDAAQAFLLSLDLRQRPMDKHYRSEVELVQVDEDDAIHVTSFPAIGGHTKPODFVILASRRKPCNGDPY
440 450 460 470 480 490 500

510 520 530 540 550 560 570
VIALRSVTLPTHRPTPEYRGGETLCSGFCILMBESDQLTKVSYNQATPGVLNVTYTVNAGLSSFFYTFKAC
VIALRSVTLPTHRPTPEYRGGETLCSGFCILMBESDQLTKVSYNQATPGVLNVTYTVNAGLSSFFYTFKAC
VIALRSVTLPTHRPTPEYRGGETLCSGFCILMBESDQLTKVSYNQATPGVLNVTYTVNAGLSSFFYTFKAC
510 520 530 540 550 560 570

580 590 X
EQFLIDNRNDLAPSLQTL
EQFLIDNRNDLAPSLQTL
EQFLIDNRNDLAPSLQTL
580 590 X

Sequence Name	Description	Length	Init. Opt. Score Score	Sig. Frame

580 590 600 610 620 630 640
TATGCAACACCATTCAGGACCTCTGCGCAACTGCGCATTCAGGCGGATCTGGAAGACAGAGACTGTAG
TATGCAACACCATTCAGGACCTCTGCGCAACTGCGCATTCAGGCGGCGATCTGGAAGACAGAGACTGTAG
TATGCAACACCATTCAGGACCTCTGCGCAACTGCGCATTCAGGCGGCGATCTGGAAGACAGAGACTGTAG
580 590 600 610 620 630 640

[illegible]

1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100	2110	2120	2130	2140	2150	2160	2170	2180	2190	2200	2210	2220	2230	2240	2250	2260	2270	2280	2290	2300	2310	2320	2330	2340	2350	2360	2370	2380	2390	2400	2410	2420	2430	2440	2450	2460	2470	2480	2490	2500	2510	2520	2530	2540	2550	2560	2570	2580	2590	2600	2610	2620	2630	2640	2650	2660	2670	2680	2690	2700	2710	2720	2730	2740	2750	2760	2770	2780	2790	2800	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900	2910	2920	2930	2940	2950	2960	2970	2980	2990	3000	3010	3020	3030	3040	3050	3060	3070	3080	3090	3100	3110	3120	3130	3140	3150	3160	3170	3180	3190	3200	3210	3220	3230	3240	3250	3260	3270	3280	3290	3300	3310	3320	3330	3340	3350	3360	3370	3380	3390	3400	3410	3420	3430	3440	3450	3460	3470	3480	3490	3500	3510	3520	3530	3540	3550	3560	3570	3580	3590	3600	3610	3620	3630	3640	3650	3660	3670	3680	3690	3700	3710	3720	3730	3740	3750	3760	3770	3780	3790	3800	3810	3820	3830	3840	3850	3860	3870	3880	3890	3900	3910	3920	3930	3940	3950	3960	3970	3980	3990	4000	4010	4020	4030	4040	4050	4060	4070	4080	4090	4100	4110	4120	4130	4140	4150	4160	4170	4180	4190	4200	4210	4220	4230	4240	4250	4260	4270	4280	4290	4300	4310	4320	4330	4340	4350	4360	4370	4380	4390	4400	4410	4420	4430	4440	4450	4460	4470	4480	4490	4500	4510	4520	4530	4540	4550	4560	4570	4580	4590	4600	4610	4620	4630	4640	4650	4660	4670	4680	4690	4700	4710	4720	4730	4740	4750	4760	4770	4780	4790	4800	4810	4820	4830	4840	4850	4860	4870	4880	4890	4900	4910	4920	4930	4940	4950	4960	4970	4980	4990	5000	5010	5020	5030	5040	5050	5060	5070	5080	5090	5100	5110	5120	5130	5140	5150	5160	5170	5180	5190	5200	5210	5220	5230	5240	5250	5260	5270	5280	5290	5300	5310	5320	5330	5340	5350	5360	5370	5380	5390	5400	5410	5420	5430	5440	5450	5460	5470	5480	5490	5500	5510	5520	5530	5540	5550	5560	5570	5580	5590	5600	5610	5620	5630	5640	5650	5660	5670	5680	5690	5700	5710	5720	5730	5740	5750	5760	5770	5780	5790	5800	5810	5820	5830
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

[illegible]

1950 1960 1970 1980

Blank Sheet

IntelIGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-10-055-624b-3.res made by tport on Tue 24 Aug 104 10:20:31-PST.

Query sequence being compared:	US-10-055-624B-3 (1-1818)
Number of sequences searched:	20
Number of scores above cutoff:	20

Results of the initial comparison of US-10-055-624B-3 (1-1818) with:
File : US10055624B.seq

STDEV	SCORE	S	C	E	R	B	M	U	N
1	606								
2	1212								
3	1818								

PARAMETERS

	Unitary	K-tuple Joining penalty Window size
Similarity matrix	1	3
Mismatch penalty	5.00	500
Gap penalty	0.33	
Gap size penalty		
Cutoff score	1	
Randomization group	0	

SEARCH STATISTICS

	Mean	Median	Standard Deviation
Scores:	229	16	547.44
Times:	CPU		Total Elapsed
	00:00:00.00		00:00:00.00

Number of residues:	6764
Number of sequences searched:	20
Number of scores above cutoff:	20

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found.

Sequence Name	Description	Init.	Opt.	Length	Score	Score	Sig.	Frame

1. US-10-055-624B-3 Sequence 3, Application U	1818	1818	1818	2.90	0
The list of other best scores is:					

The list of other best scores is:

Sequence Name	Description	Length	Score	Score Sig.	Frame
---------------	-------------	--------	-------	------------	-------

	*** 2 standard deviations above mean ***		
2. US-10-055-624B-1 Sequence 1, Application U	1857	1633	1692
1. US-10-055-624B-1 standard deviation above mean	***		
3. US-10-055-624B-5 Sequence 5, Application U	2699	876	1529
			1.18

1. US-10-055-624B-3 (1-1818)
US-10-055-624B-3 Sequence 3, Application US/10055624E

```
Initial Score = 1818 Optimized Score = 1818 Significance = 2.90
Residue Identity = 100% Matches = 1818 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0
```

X 10 20 30 40 50 60 70
GGGAAATGGCCCTTGCATCAAAATGATCCGAATGTGGAAATCACCTGCGACGGGCTTGGCCCTCTGTGTT
GGGAAATGGCCCTTGCATCAAAATGATCCGAATGTGGAAATCACCTGCGACGGGCTTGGCCCTCTGTGTT
GGGAAATGGCCCTTGCATCAAAATGATCCGAATGTGGAAATCACCTGCGACGGGCTTGGCCCTCTGTGTT
GGGAAATGGCCCTTGCATCAAAATGATCCGAATGTGGAAATCACCTGCGACGGGCTTGGCCCTCTGTGTT
X 10 20 30 40 50 60 70

80 90 100 110 120 130 140
CTTCAACCGCACATCCCGGAAGTACAGCTTACGTGCGGGAAAGATGCGGACATGCGACGCGGACGAGATAT
CTTCAACCGCACATCCCGGAAGTACAGCTTACGTGCGGGAAAGATGCGGACATGCGACGCGGACGAGATAT
80 90 100 110 120 130 140
CTTCAACCGCACATCCCGGAAGTACAGCTTACGTGCGGGAAAGATGCGGACATGCGACGCGGACGAGATAT

150 160 170 180 190 200 210
 CCGGAACCCACGGAGGTGCGAGTAGCGCACTGTGTCTCCCTGCCACCAACCAAGTGTAGCTGAA
 CCGGAACCCACGGAGGTGCGAGTAGCGCACTGTGTCTCCCTGCCACCAACCAAGTGTAGCTGAA
 CCGGAACCCACGGAGGTGCGAGTAGCGCACTGTGTCTCCCTGCCACCAACCAAGTGTAGCTGAA
 150 160 170 180 190 200 210

220 230 240 250 260 270 280
CGTCGGGAGCGCTCAAGTGATTCACCA CGGCTTGCTGTCCGGAGAGGCACGCTGGCTGCCCT
CGTCGGGAGCGCTCAAGTGATTCACCA CGGCTTGCTGTCCGGAGAGGCACGCTGGCTGCCCT
CGTCGGGAGCGCTCAAGTGATTCACCA CGGCTTGCTGTCCGGAGAGGCACGCTGGCTGCCCT
220 230 240 250 260 270 280

290 300 310 320 330 340 350 360
TGTCAAGCTTCATGATGACATATTGTCAGACACCATTTGTTGACAAGTGGGATATCAAGAGG
TGTCAAGCTTCATGATGACATATTGTCAGACACCATTTGTTGACAAGTGGGATATCAAGAGG
TGTCAAGCTTCATGATGACATATTGTCAGACACCATTTGTTGACAAGTGGGATATCAAGAGG
TGTCAAGCTTCATGATGACATATTGTCAGACACCATTTGTTGACAAGTGGGATATCAAGAGG

370 380 390 400 410 420 430
CAAGTGAACCCGGGCTTCAATCCAGGATGGAGTGGCATCCAGGTGGCTCGAGAGCACTGTGCTGG
CAAGTGAACCCGGGCTTCAATCCAGGATGGAGTGGCATCCAGGTGGCTCGAGAGCACTGTGCTGG
CAAGTGAACCCGGGCTTCAATCCAGGATGGAGTGGCATCCAGGTGGCTCGAGAGCACTGTGCTGG
370 380 390 400 410 420 430

440
450
460
470
480
490
500

GAAGCAGTGAATGTGTGAAGGCCCTTGGCCCACTTCTGTGGCCCGCGAGATCACCAAGTGAACCTGA
GAAGCAGTGAATGTGTGAAGGCCCTTGGCCCACTTCTGTGGCCCGCGAGATCACCAAGTGAACCTGA
GAAGCAGTGAATGTGTGAAGGCCCTTGGCCCACTTCTGTGGCCCGCGAGATCACCAAGTGAACCTGA

GAGATACGCGCCGACAGAGAGAGATGTGGGCGCTTGACCGCGGCGCATTCGGC
 510 520 530 540 550 560 570
 GCGATACCGCCGACAGAGAGAGATGTGGGCGCTTGACCGCGGCGCATTCGGC
 510 520 530 540 550 560 570

580	590	600	610	620	630	640
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGACAGAGACT						
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGACAGAGACT						
TGTTATGACAGACCATCAAGACCTCTGGCCAACTGGCCATTGAGGCGATCTGGACAGAGACT						
580	590	600	610	620	630	640

650	660	670	680	690	700	710	720
TAGCCGATGTGTCGGGTTGAAGACCCGTTGAGAGTGTGAGTGTGAGTGTCTGCTCCGCCACCAATATC							
TAGCCGATGTGTCGGGTTGAAGACCCGTTGAGAGTGTGAGTGTGAGTGTCTGCTCCGCCACCAATATC							
650	660	670	680	690	700	710	720
CCAGGGCAACACCTTTGGGGGGCCAGATCATGCGCTGATGGAATGTGGCCCAATTCAGACAGCCGGCT							
CCAGGGCAACACCTTTGGGGGGCCAGATCATGCGCTGATGGAATGTGGCCCAATTCAGACAGCCGGCT							
730	740	750	760	770	780	790	
CTGCCGATGCCACCCCTGCGCTGAAGGCCATTTGAATGTTCCACTTCCGAGGCCGCTCCAGATCGGCGACCG							
CTGCCGATGCCACCCCTGCGCTGAAGGCCATTTGAATGTTCCACTTCCGAGGCCGCTCCAGATCGGCGACCG							
800	810	820	830	840	850	860	
CTGCCGATGCCACCCCTGCGCTGAAGGCCATTTGAATGTTCCACTTCCGAGGCCGCTCCAGATCGGCGACCG							
CTGCCGATGCCACCCCTGCGCTGAAGGCCATTTGAATGTTCCACTTCCGAGGCCGCTCCAGATCGGCGACCG							
800	810	820	830	840	850	860	
TTGTGTCTCAAAACCTCATCGTGAACAATGCTTCACAAATAGATGAGAGTGGGCTGTGCTGAGAGCCCTTA							
TTGTGTCTCAAAACCTCATCGTGAACAATGCTTCACAAATAGATGAGAGTGGGCTGTGCTGAGAGCCCTTA							
870	880	890	900	910	920	930	
TTGTGTCTCAAAACCTCATCGTGAACAATGCTTCACAAATAGATGAGAGTGGGCTGTGCTGAGAGCCCTTA							
TTGTGTCTCAAAACCTCATCGTGAACAATGCTTCACAAATAGATGAGAGTGGGCTGTGCTGAGAGCCCTTA							
870	880	890	900	910	920	930	
TGCGCAGAGGCTGTAGACCCAGCGGCCACATCAACAGTGGCTTATGACCTTTGTGTCTTGAAGCGACA							
TGCGCAGAGGCTGTAGACCCAGCGGCCACATCAACAGTGGCTTATGACCTTTGTGTCTTGAAGCGACA							
940	950	960	970	980	990	1000	
TGCGCAGAGGCTGTAGACCCAGCGGCCACATCAACAGTGGCTTATGACCTTTGTGTCTTGAAGCGACA							
TGCGCAGAGGCTGTAGACCCAGCGGCCACATCAACAGTGGCTTATGACCTTTGTGTCTTGAAGCGACA							
940	950	960	970	980	990	1000	
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
TGACCAAGCCCAAGTGTGCTGCTCGGATTTGGGCCCAAGCCCGCGCATGTGAGCGAGGCTTCCAGAGGCCAG							
1010	1020	1030	1040	1050	1060	1070	1080
TG							

[illegible]

440 450 460 470 480 490 500
GAAGCAGTGGAAATGTGTGCAAGGCTTTGACCTTCGTGTCGCCCCGCTCGAAGATACCAAGTGAAGCTTGA
GAAGCAGTGGAAATGTGTGCAAGGCTTTGACCTTCGTGTCGCCCCGCTCGAAGATACCAAGTGAAGCTTGA
430 440 450 460 470 480 490 500
510 520 530 540 550 560 570
GCAGATCACGCGCGCGGACGAAAGAGGAAGATGAGCACAAGTGTGGCGCTTGAAGCGCCGTCGATGTGGCT
GCAGATCACGCGCGCGGACGAAAGAGGAAGATGAGCACAAGTGTGGCGCTTGAAGCGCCGTCGATGTGGCT
580 590 600 610 620 630 640
TGTCTATGCAAGACCATCAAGAGACCTCTGTGGCCAACTGTGGCCATTTCAGGGGATGTGAGAGCAGACTG
TGTCTATGCAAGACCATCAAGAGACCTCTGTGGCCAACTGTGGCCATTTCAGGGGATGTGAGAGCAGACTG
580 590 600 610 620 630 640
TAGCCGCGATGTGCGCGCTGAGAAAGACCGCGTGTGAGAGTGTGAGATGTGCTCTGCTCCACGCAATCA
TAGCCGCGATGTGCGCGCTGAGAAAGACCGCGTGTGAGAGTGTGAGATGTGCTCTGCTCCACGCAATCA
650 660 670 680 690 700 710 720
CCAGAGGAAACACCTTTGGGGGCGCAGATCATGTGCTGTGAGATGAGATGTGGCACAATTGTGACGACCGGCT
CCAGAGGAAACACCTTTGGGGGCGCAGATCATGTGCTGTGAGATGAGATGTGGCACAATTGTGACGACCGGCT
730 740 750 760 770 780 790
CTGCGCGCCACCACTTACGCTGAGAGGCGCATTAAGTATTCACCTTCGAGAGCCGCTCCAGATGGCGACCG
CTGCGCGCCACCACTTACGCTGAGAGGCGCATTAAGTATTCACCTTCGAGAGCCGCTCCAGATGGCGACCG
800 810 820 830 840 850 860
CTGCGCGCCACCACTTACGCTGAGAGGCGCATTAAGTATTCACCTTCGAGAGCCGCTCCAGATGGCGACCG
CTGCGCGCCACCACTTACGCTGAGAGGCGCATTAAGTATTCACCTTCGAGAGCCGCTCCAGATGGCGACCG
790 800 810 820 830 840 850 860
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
870 880 890 900 910 920 930
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
870 880 890 900 910 920 930
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
TCTGTGTCTCAAAAGCCATCGGAACAAATGCTTCAAAATATGATGATGAGAGTGGCGGTGCGTGGAGAGGCTTA
940 950 960 970 980 990 1000
TCGCGAGGAGGCTGAGACCCACCGCGCGCAATACAGATGTCCTTTATGACCTTTGTGTCTTGAACGACGA
TCGCGAGGAGGCTGAGACCCACCGCGCGCAATACAGATGTCCTTTATGACCTTTGTGTCTTGAACGACGA
940 950 960 970 980 990 1000
TCGCGAGGAGGCTGAGACCCACCGCGCGCAATACAGATGTCCTTTATGACCTTTGTGTCTTGAACGACGA
TCGCGAGGAGGCTGAGACCCACCGCGCGCAATACAGATGTCCTTTATGACCTTTGTGTCTTGAACGACGA
1010 1020 1030 1040 1050 1060 1070 1080
TGACGAGACCCAGTTGCTGTCGCTGTGATTCGCGCCAGCCGCGCATGTGTGACGCGGTATCCGAGAGGCTGAG
TGACGAGACCCAGTTGCTGTCGCTGTGATTCGCGCCAGCCGCGCATGTGTGACGCGGTATCCGAGAGGCTGAG
1010 1020 1030 1040 1050 1060 1070
TGACGAGACCCAGTTGCTGTCGCTGTGATTCGCGCCAGCCGCGCATGTGTGACGCGGTATCCGAGAGGCTGAG
TGACGAGACCCAGTTGCTGTCGCTGTGATTCGCGCCAGCCGCGCATGTGTGACGCGGTATCCGAGAGGCTGAG
1090 1100 1110 1120 1130 1140 1150
TGCGCAAGAAAGATCCGCTTGAACAGAAAGTACATCGTCTCTGTATAGCAGACAGAGTCCCCCTTCGCT
TGCGCAAGAAAGATCCGCTTGAACAGAAAGTACATCGTCTCTGTATAGCAGACAGAGTCCCCCTTCGCT
1080 1090 1100 1110 1120 1130 1140
TGCGCAAGAAAGATCCGCTTGAACAGAAAGTACATCGTCTCTGTATAGCAGACAGAGTCCCCCTTCGCT
TGCGCAAGAAAGATCCGCTTGAACAGAAAGTACATCGTCTCTGTATAGCAGACAGAGTCCCCCTTCGCT
1150 1160 1170 1180 1190 1200 1210 1220
CCCCCGGAGCCCTAGGAAACAGAGTATACCTGAGCTACAAATAGCTCTCTCTTGAGAAATGCTTGTGGCAAA
CCCCCGGAGCCCTAGGAAACAGAGTATACCTGAGCTACAAATAGCTCTCTCTTGAGAAATGCTTGTGGCAAA
1150 1160 1170 1180 1190 1200 1210 1220
CCCCCGGAGCCCTAGGAAACAGAGTATACCTGAGCTACAAATAGCTCTCTCTTGAGAAATGCTTGTGGCAAA
CCCCCGGAGCCCTAGGAAACAGAGTATACCTGAGCTACAAATAGCTCTCTCTTGAGAAATGCTTGTGGCAAA
1230 1240 1250 1260 1270 1280 1290
GGCACAATCGGTGTCTCCGAGATCACTGATCAGTCCGCTGTACCTCTGGAAGATGACAAGTTCTCTCTC
GGCACAATCGGTGTCTCCGAGATCACTGATCAGTCCGCTGTACCTCTGGAAGATGACAAGTTCTCTCTC
1230 1240 1250 1260 1270 1280 1290
GGCACAATCGGTGTCTCCGAGATCACTGATCAGTCCGCTGTACCTCTGGAAGATGACAAGTTCTCTCTC
GGCACAATCGGTGTCTCCGAGATCACTGATCAGTCCGCTGTACCTCTGGAAGATGACAAGTTCTCTCTC

[illegible]

[illegible]

1030	1040	1050	1060	1070	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
TGCTGCTGAGATTCGGGCGGCGCCGACATGGTTGAGCGCGGTACCTCGAGGCGTCAATGTCGCAAGAAAGAG	ACTGCGCTGGATTCGTCCTCCAGCTCTGGAGGGGTGAACGGCATTCGACAGGACCGAGGAAAGAG	1080	1090	1100	1110	1120	1130	1140	1150	1160	1170	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000			
TCGGCTCTGAGACAGGAATACATCCTGTCTCTGTATGACGACAGACAGGTGCGCCCTCTCCCTCCCTGGACCTT	TCGGCTGACAGGAAATACCTTGTCTCTGTATGACGACAGGACAGGTGCGCCCTCTCCCTCCCTGGACCTT	1150	1160	1170	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000										
GCAACACAGGTATACCTGAGCTACCAATACCTCTCTCTTGAAGATGTTGTGGGCCAAGAGCAACTGGGTGCG	GCAACACAGGTATACCTGAGCTACCAATACCTCTCTCTTGAAGATGTTGTGGGCCAAGAGCAACTGGGTGCG	1170	1180	1190	1200	1210	1220	1230	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740	1750	1760	1770	1780	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000												
TGTCTCTGAGATTCAGTCAAGTCCGCTGCTGTACATCTGTGAGAGATGACAAGATTCCTCTCTCTTACATGAGAG	TGTCTCTGAGATTCAGTCAAGTCCGCTGCTGTACATCTGTGAGAGATGACAAGATTCCTCTCTCTTACATGAGAG	1240	1250	1260	1270	1280	1290	1300	1310	1320	1330	1340	1350	1360	1370	1380	1390	1400	1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	1710	1720	1730	1740																																													

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:45:27 ; Search time 49 Seconds
(without alignments)
3888.853 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181
Sequence: 1 MTQNVGNHLRRGLASVFSNR.....TLKAIIVAKINSRFGYLDPT 607

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1292805 seqs, 313927144 residues

Total number of hits satisfying chosen parameters: 1292805

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3181	100.0	607	US-10-055-624B-2	Sequence 2, Appl1
2	2838	89.2	594	US-10-055-624B-4	Sequence 4, Appl1
3	2530	79.5	594	US-10-055-624B-6	Sequence 6, Appl1
4	1242.5	39.1	555	US-10-336-472-50	Sequence 59, Appl1
5	413	13.0	119	US-09-925-298-599	Sequence 599, App
6	413	13.0	119	US-10-102-806-599	Sequence 599, App
7	255	8.0	339	US-09-738-626-6344	Sequence 6344, App
8	235	8.0	339	US-10-627-476-230	Sequence 230, App
9	223.5	7.0	155	US-10-282-122A-54710	Sequence 54710, A
10	223.5	7.0	155	US-10-289-762-693	Sequence 693, App
11	220	6.9	339	US-09-809-391-488	Sequence 488, App
12	220	6.9	339	US-09-882-171-488	Sequence 488, App
13	220	6.9	339	US-10-164-861-488	Sequence 488, App
14	219	6.9	370	US-10-094-749-2967	Sequence 2967, App
15	207.5	6.5	159	US-10-282-122A-55313	Sequence 55313, A

15	203	6.4	168	12	US-10-282-122A-45753	Sequence 45753, A
17	201	6.3	207	14	US-10-156-761-12563	Sequence 12563, A
18	186	5.8	156	12	US-10-282-122A-45621	Sequence 45621, A
19	185	5.8	176	12	US-10-282-122A-70623	Sequence 70623, A
20	182	5.7	455	10	US-09-906-408A-8	Sequence 8, Appl1
21	181.5	5.6	183	12	US-10-282-122A-63369	Sequence 63369, A
22	177	5.6	174	12	US-10-282-122A-58829	Sequence 58829, A
23	177	5.6	174	12	US-10-335-977-8306	Sequence 8306, App
24	177	5.6	177	12	US-10-335-977-8307	Sequence 8307, App
25	173	5.4	172	12	US-10-282-122A-60467	Sequence 60467, A
26	171.5	5.4	170	12	US-10-282-122A-45497	Sequence 45497, A
27	171.5	5.4	177	12	US-10-282-122A-71369	Sequence 71369, A
28	171.5	5.4	526	16	US-10-437-963-119373	Sequence 119373, A
29	170.5	5.4	176	12	US-10-282-122A-70386	Sequence 70386, A
30	166	5.2	178	12	US-10-282-122A-56911	Sequence 56911, A
31	159.5	5.0	434	12	US-10-424-599-151287	Sequence 151287, A
32	146.5	4.6	321	12	US-10-092-900A-222	Sequence 222, App
33	140.5	4.4	110	12	US-10-424-599-220213	Sequence 220213, A
34	140.5	4.4	438	10	US-09-906-408A-7	Sequence 7, Appl1
35	140	4.4	137	12	US-10-282-122A-50648	Sequence 50648, A
36	139.5	4.4	131	12	US-10-282-122A-49795	Sequence 49795, A
37	133	4.2	137	12	US-10-282-122A-54459	Sequence 54459, A
38	131	4.0	136	16	US-10-408-765A-2655	Sequence 2655, A
39	128	4.0	136	12	US-10-282-122A-51198	Sequence 51198, A
40	124.5	3.9	136	12	US-10-282-122A-48235	Sequence 48235, A
41	121	3.8	2108	8	US-08-973-363-16	Sequence 16, Appl1
42	119.5	3.8	140	12	US-10-282-122A-44586	Sequence 44586, A
43	118	3.7	445	14	US-10-177-293-296	Sequence 296, App
44	118	3.7	445	15	US-10-435-696-33	Sequence 33, Appl1
45	115	3.6	1289	9	US-09-932-145-11	Sequence 11, Appl1

ALIGNMENTS

RESULT 1
US-10-055-624B-2
; Sequence 2, Application US/10055624B
; Publication No. US20030220238A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Charles A
; APPLICANT: Goddard, Audrey D
; APPLICANT: Grimaldi, J Christopher
; TITLE OF INVENTION: BFT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-624B-2

Query Match	Similarity	Score	DB	Length	607;
Best Local	100.0%	Pred. No. 0;			
Matches	607;	Conservative	0;	Mismatches	0;
QY	1	MTQNVGNHLRRGLASVFSNR	1	MTQNVGNHLRRGLASVFSNR	60
DB	1	MTQNVGNHLRRGLASVFSNR	1	MTQNVGNHLRRGLASVFSNR	60
QY	61	RGEISVGQLKXIDTTACLSAERRAGCPVTASMDIYFEHTISVGQVNNIKAKNNRAFN	120		
DB	61	RGEISVGQLKXIDTTACLSAERRAGCPVTASMDIYFEHTISVGQVNNIKAKNNRAFN	120		
QY	121	SSMEVGIOVASBDDCSEKQWVCKALATFVARREITTKKQITRTEEBKXEHSAABR	180		
DB	121	SSMEVGIOVASBDDCSEKQWVCKALATFVARREITTKKQITRTEEBKXEHSAABR	180		

```

QY 181 RRMRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEFG 240
Db 181 RRMRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEFG 240
QY 241 QIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEVG 300
Db 241 QIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEVG 300
QY 301 CVEAAROEAEFTHRRHNSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
Db 301 CVEAAROEAEFTHRRHNSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
QY 361 DRKYIVSCKQTEVPLSVMPDPSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLED 420
Db 361 DRKYIVSCKQTEVPLSVMPDPSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLED 420
QY 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
Db 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
QY 481 GHTKPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRRGETLCSGFCLMRBGDQ 540
Db 481 GHTKPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRRGETLCSGFCLMRBGDQ 540
QY 541 LTKCCWVRVSLTEIVSASGFYSNGLBSRSKGRSDGMNGKLAGHLSITLKAIPVAKINSR 600
Db 541 LTKCCWVRVSLTEIVSASGFYSNGLBSRSKGRSDGMNGKLAGHLSITLKAIPVAKINSR 600
QY 601 FGYLQDT 607
Db 601 FGYLQDT 607

```

RESULT 2

```

US-10-055-624B-4
; Sequence 4, Application US/10055624B
; Publication No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; TITLE OF INVENTION: BIT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-055-624B-4

```

```

Query Match      89.2% Score 2838; DB 15; Length 594;
Best Local Similarity 100.0%; Pred. No. 2e-283;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 60
Db 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 60
QY 61 RGEISVQQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFNF 120
Db 61 RGEISVQQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFNF 120
QY 121 SSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 180
Db 121 SSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 180

```

```

QY 181 RRMRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEFG 240
Db 181 RRMRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEFG 240
QY 241 QIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEVG 300
Db 241 QIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEVG 300
QY 301 CVEAAROEAEFTHRRHNSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
Db 301 CVEAAROEAEFTHRRHNSAFMTFVVLADDDQPOLPMIRPOGDDGRRRREASARKKIRL 360
QY 361 DRKYIVSCKQTEVPLSVMPDPSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLED 420
Db 361 DRKYIVSCKQTEVPLSVMPDPSNQYVLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLED 420
QY 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
Db 421 DKFLSFHMEMVHVDAQAQAFLLISDLRORPEMDKHYRSVELVQVDEDDAIYHVTSPALG 480
QY 481 GHTKPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRRGETLCSGFCLMRBGDQ 540
Db 481 GHTKPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRRGETLCSGFCLMRBGDQ 540
QY 541 LTK 543
Db 541 LTK 543

```

RESULT 3

```

US-10-055-624B-6
; Sequence 6, Application US/10055624B
; Publication No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; TITLE OF INVENTION: BIT COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055,624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-055-624B-6

```

```

Query Match      79.5% Score 2530; DB 15; Length 594;
Best Local Similarity 81.9%; Pred. No. 1.3e-251;
Matches 493; Conservative 35; Mismatches 54; Indels 20; Gaps 6;

```

```

QY 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 58
Db 1 MIQVGNHILRGGLASVFSNRTSRKSAIRAGNDSAMADGEGYRNPTVOMSOVLPCHTNQ 58
QY 59 NQRELSIGQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFNF 118
Db 59 NQRELSIGQLIKMIDTTACLSAERHAGCPCVTASMDIYFEHTISVGVNNIKAVNAFNF 118
QY 119 FNSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 178
Db 119 FNSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 178
QY 121 FNSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 180
Db 121 FNSMEVGIQVASEDLCSKQNNVCKALATFVARREITVKLKOITPRTEEEKHESVAER 180
QY 179 ERRMRRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEF 238
Db 179 ERRMRRLVYADTITKDLANCAIQGDLBSRDCSRWPAKTRVESVELVLPHPANHQNTEF 238
QY 239 GQIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEV 298
Db 239 GQIMAMMENAVATIAASRLCAHPTLKAIEHFHRRGSPQVGDRLVLAIVNNAFKHSMEV 298

```

```
Db      240 GGGIMAMMENAVATIAASRLCHAPTLKALIMFHFRRPSQVGDVLKAIIVNAFKHSMEX 299
Qy      299 GVCVEAYROEATHRHINSAPMTFVYLDADDPQLLPWIRPQDGERRYRASARKKI 358
Db      300 GVCVEAYROEATHRHINSAPMTFVYLDADDPQLLPWIRPQDGERRYRASARKKI 359
Qy      359 RLDKRYIVSCQTEVPLVSPWMDPSNOVYLSYNNVSSIKMLVAKDNVLSSEISQVRLYTL 418
Db      360 RLDKRYIVSCQTEVPLVSPWMDPSNOVYLSYNNVSSIKMLVAKDNVLSSEISQVRLYTL 419
Qy      419 EDKPEISFHEMRYVYDAQAFLILSDLRORPEWMDKHYNSVEIVQVDEDDAIYHYTSPA 478
Db      420 EED-FLSPHLEWVYVDAQVQLLSDLRORPEWMDKHYNSVEIVQVDEDDAIYHYTSPA 478
Qy      479 LGGHTPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPERRGETLCSGFCIMRBS 538
Db      479 LSGNTKPODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPERRGETLCSGFCIMRBS 538
Qy      539 DQITKCVSYNN-----QATPGFLNY-VTINVSGLSSEFYN-----TFKACSEFLLD 582
Db      539 SR 600
Qy      583 NR 584
```

RESULT 4

```
US-10-336-472-50
; Sequence 50, Application US/10336472
; Publication No. US20040043929A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Anderson, David W.
; APPLICANT: Ballinger, Robert A.
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Chant, John S.
; APPLICANT: Berghs, Constance
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gilbert, Jennifer A.
; APPLICANT: Gunther, Erik
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Li, Li
; APPLICANT: Liu, Xiaohong
; APPLICANT: Miller, Charles E.
; APPLICANT: Miller, Isabelle
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Rastelli, Luca
; APPLICANT: MacDougall, John R.
; APPLICANT: Mishra, Vishnu
; APPLICANT: Penz, Carol E.A.
; APPLICANT: Spaderina, Steven K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Ort, Taciana
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Tohernev, Velizar T.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Zernhusen, Bryan D.
; APPLICANT: Zhong, Mei
```

```
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-533C
; CURRENT APPLICATION NUMBER: US/10/336,472
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuroseqList version 0.1
; SEQ ID NO 50
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-50
```

```
Query Match      39.1%; Score 1242.5; DB 12; Length 555;
Best Local Similarity 52.2%; Pred. No. 1.2e-118;
Matches 260; Conservative 82; Mismatches 143; Indels 13; Gaps 7;
```

```
Qy      44 PTEVQMSQVLPCHTNORGELSVGQLKWIIDTACLSARHAGCPVTASMDIYENT 103
Db      6 PGEVVMQSAIOPRAHNAARELSAGQLKWIIDTACLSARHAGCPVTASMDIYENT 65
Qy      104 SVGQVNIYAKVNAFNSMVEGIOVASDLCSEKQWNYCKALATVAV--REITKVL 160
Db      66 RVGQVITIKAKYTRAFSTMEISIKVMQDMLTGIEKIVSVAFSTVAKVGKE--KIHL 123
Qy      161 KQITPREBKKHSHVAARBRMRIVYADTIDOLANCAIQGLSRRDSCRMVPAEKTRY 220
Db      124 KPVTLILBQDHEBHNAARERKVRLOHEDTFNNLMKSSKFDLLIFDEEGAVSTGTSV 183
Qy      221 ESVEIVLPFHANHQNTFGQIMAMMENAVATIAASRLCHAPTLKALIMFHFRRPSQVGD 280
Db      184 QGIEIVLPFHANHHNGTFFGQIMAMMETVATISASLCLMHPRLKSVDMKFGPSTVGD 243
Qy      281 RLVKAIYVNAFKHNSVEVGVCEAY--ROEATHRHINSAPMTFVYLDADDPQL--P 336
Db      244 RLVFTAIYVNTFQTCVEGVVEAFDQCEMARGRGHINSAPF--IYNSADDEKENLITFP 301
Qy      337 WIRPQDGERRRRRESARKKIRLDKRYIVSCQTEVPLVSPWMDPSNOVYLSYNNVSLK 396
Db      302 RIQPLSKDPPRRKRGAIARKRIRLGRKVIYSHK--EVLICIMWDISKQSLSSNVAALK 360
Qy      397 MIVAKDNVLSSEISQVRLYTTLEDKELSFHEMRYVYDAQAFLILSDLRORPEWMDKHY 456
Db      361 KLAARKGMEVTSIVKEIKITYLIEHDVLISVMVEKHVGSRAHLAYRLSDPTKRLMDPHF 420
Qy      457 KSVELVQVDEDDAIYHYTSPAIGHTXQDFVILASRRKPCDNGDPYVIALRSVTLPTH 516
Db      421 VSCVEVIDWVSEDDQHLHITCPL--NDCKPKDLVILVSRKPKLDKQNTYVAVXSVLLPSV 479
Qy      517 RETPEYRGETLCSGFCIL 534
Db      480 PPSQYITRSEITCAGFLI 497
```

```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 NAME/KEY: SITE
4 LOCATION: (58)
5 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
6 NAME/KEY: SITE
7 LOCATION: (68)
8 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
9 NAME/KEY: SITE
10 LOCATION: (88)
11 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
12 NAME/KEY: SITE
13 LOCATION: (98)
14 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
15 NAME/KEY: SITE
16 LOCATION: (99)
17 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
18 US-10-102-806-599
19
20 Query Match 13.0%; Score 413; DB 14; Length 119;
21 Best Local Similarity 96.4%; Pred. No. 4.6e-34;
22 Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
23
24 QY 383 NYYLTVNNVSSLSKMYIAKNWVYSSRISQVRLTYTDEDKFLSPHMMVYVVAQAQFL 442
25 DB 4 DQVLSLNNVSSLSKMLVAKNNWVYSSRISQVRLTYTDEDKFLSPHMMVYVVAQAQFL 63
26
27 QY 443 LSDLRQRPENDKHYRSVELVQOVD 466
28 DB 64 LSDLRQRPENDKHYRSVELVQOVD 87
29
30 RESULT 7
31 US-09-738-626-6344
32 Sequence 6344, Application US/09738626
33 Publication No. US20020197605A1
34 GENERAL INFORMATION:
35 APPLICANT: NAKAGAWA, SATOSHI
36 APPLICANT: MIZOGUCHI, HIROSHI
37 APPLICANT: ANDO, SEIKO
38 APPLICANT: HAYASHI, MIKIRO
39 APPLICANT: OCHIAI, KEIKO
40 APPLICANT: YOKOI, HARUHIKO
41 APPLICANT: TATEISHI, NAOKO
42 APPLICANT: SENOH, AKIHIRO
43 APPLICANT: IKEDA, MASATO
44 APPLICANT: OZAKI, AKIO
45 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
46 FILE REFERENCE: 249-125
47 CURRENT APPLICATION NUMBER: US/09/738, 626
48 PRIOR FILING DATE: 2000-12-18
49 PRIOR APPLICATION NUMBER: JP 99/377484
50 PRIOR FILING DATE: 1999-12-16
51 PRIOR APPLICATION NUMBER: JP 00/159162
52 PRIOR FILING DATE: 2000-04-07
53 PRIOR APPLICATION NUMBER: JP 00/280988
54 PRIOR FILING DATE: 2000-08-03
55 NUMBER OF SEQ ID NOS: 7059
56 SOFTWARE: PatentIn ver. 3.0
57 SEQ ID NO 6344
58 LENGTH: 339
59 TYPE: PRT
60 ORGANISM: Corynebacterium glutamicum
61 US-09-738-626-6344
62
63 Query Match 8.0%; Score 255; DB 9; Length 339;
64 Best Local Similarity 25.7%; Pred. No. 5.3e-17;
65 Matches 89; Conservative 56; Mismatches 17; Indels 30; Gaps 10;
66
67 QY 41 YRNPTVQMSQLVLPCHTNGRGELSY--GGLTKWIDTTACTASPRHAGCPCTASMDIY 98
68 DB 14 HRSP-EVYLRFMAAPFTVLMAGSHGVGGKVLKMDIAAAYACAIGMSGTYCVTAIVGHII 72

```

QY 99 FEHTIVGVGVNIAKAVNRANFNSSMEVGLQVASELCEKQMNVCALATVAREIT-- 156
Db 73 FTPIPSGHMVEVRSIAMTGRSSMHINEVLSAPBDGNVTRACDCLVFVADATATGR 132
QY 157 KVKLKQITPRTEEEKKSHVAAERRRMLVYADTIKDLANCAIQDDESRDCSRMPAE 216
Db 133 ATVPSEFTPKNEBQGVLEAANSRLGLAK-----ALBAEMEKOTYNG--PSE 177
QY 217 KTRVSEVELVPRPHANHGNTTGGQIYAMENAVATIAASRLCRAHPTILAKEMHFPGPS 276
Db 178 APRLLITRFLAKPTDINMGXKXHGATNEMIEBAGACTEMSGNHTVAVYAGIRFYQPI 23
QY 277 QVSDRLVLTALAIYNNAFKISMVEGVV---EAYRQEAETHRR-HINSAPMTFVLADADQP 332
Db 238 QISDLEIVARMMRTDKSSMOWSIHVRAAGDHRGALETAIHH--ATVTLGIDVDGP 294
QY 333 QLLPWIRPQGDGERRYREASARKKIRIDRXYI-VSCQTEVPLSV 377
Db 295 LPAQCFVPTPEDIOIAEHANIRILRIAD--YTMPYFORRYPIOT 338

```

US-10-282-122A-54710
/ Sequence 54710, Application US/10282122A
/ Publication No. US20040029129A1
/
GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyth, R.
/ APPLICANT: Xu, H.
/
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/
FILE REFERENCE: ELITRA.034A
/
CURRENT FILING DATE: US/10/282,122A
/
PRIOR APPLICATION NUMBER: 2003-02-20
/
PRIOR FILING DATE: 60/151,078
/
PRIOR APPLICATION NUMBER: 60/206,948
/
PRIOR FILING DATE: 2000-03-21
/
PRIOR APPLICATION NUMBER: 60/05-23
/
PRIOR FILING DATE: 2000-05-23
/
PRIOR APPLICATION NUMBER: 60/207,727
/
PRIOR FILING DATE: 2000-05-26
/
PRIOR APPLICATION NUMBER: 60/220,335
/
PRIOR FILING DATE: 2000-09-06
/
PRIOR APPLICATION NUMBER: 60/230,347
/
PRIOR FILING DATE: 2000-09-09
/
PRIOR APPLICATION NUMBER: 60/242,578
/
PRIOR FILING DATE: 2000-10-23
/
PRIOR APPLICATION NUMBER: 60/253,625
/
PRIOR FILING DATE: 2000-11-27
/
PRIOR APPLICATION NUMBER: 60/257,931
/
PRIOR FILING DATE: 2000-12-22
/
PRIOR APPLICATION NUMBER: 60/267,636
/
PRIOR FILING DATE: 2001-02-09
/
PRIOR APPLICATION NUMBER: 60/269,308
/
PRIOR FILING DATE: 2001-02-16
/
/
Remaining prior Application data removed - See File Wrapper or PAM.
/

```

```

/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO: 54710
/ LENGTH: 155
/ TYPE: PRT
/ ORGANISM: Chlamydia pneumoniae
/ JUS-10-282-122A-54710

Query Match      7.0%      Score 223.5, DB 12, Length 155;
Best Local Similarity 35.5%;      Pred. No. 2.7e-14;

```

```

Query Match Similarity 7.0%; Score 223.5; DB 15; Length 155;
Best Local Similarity 35.5%; Pred. No. 2.7e-14;
Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;

QY      226  VI.PPHAHQGTFFGQIIMAMNENNVATIAASRLCRAHPFLKALIEHFHFGSPQSGVEDRLVLK 285
      19  IFFNDLANNNTVFGGILMSLIDRLALVVAERTSEVCYTAFFVALRFAPAYMGENILCK 78
QY      286  AIYVNAAFHSHMEVGVCEVA---YRQEAETHRRHINSAPMTFVLDDADPOQLLPIRPOP 342
      79  AAVNTWTWTSLEVGKWAENITYKOE---RRHITSYIFTVAVAVEDNQPLFVHQIVETF 134
Db      343  GDGERRYREASARKKIRLDK 363
      135  PEKKRRYNEADRRRQARLEK 155

RESULT 11
US-09-809-391-488
; Sequence 488, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 488
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (142)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

```

! PRIOR FILING DATE: 19900500
! PRIOR APPLICATION NUMBER: 60/047,600
;

1	PRIOR FILING DATE: 1997-05-23	1	PRIOR APPLICATION NUMBER: 60/047,615
2	PRIOR FILING DATE: 1997-05-23	2	PRIOR FILING DATE: 1997-08-22
3	PRIOR APPLICATION NUMBER: 60/047,597	3	PRIOR APPLICATION NUMBER: 60/056,893
4	PRIOR FILING DATE: 1997-05-23	4	PRIOR FILING DATE: 1997-08-22
5	PRIOR APPLICATION NUMBER: 60/047,502	5	PRIOR APPLICATION NUMBER: 60/056,633
6	PRIOR FILING DATE: 1997-05-23	6	PRIOR FILING DATE: 1997-08-22
7	PRIOR APPLICATION NUMBER: 60/047,633	7	PRIOR APPLICATION NUMBER: 60/056,878
8	PRIOR FILING DATE: 1997-05-23	8	PRIOR FILING DATE: 1997-08-22
9	PRIOR APPLICATION NUMBER: 60/047,583	9	PRIOR APPLICATION NUMBER: 60/056,662
10	PRIOR FILING DATE: 1997-05-23	10	PRIOR FILING DATE: 1997-08-22
11	PRIOR APPLICATION NUMBER: 60/047,617	11	PRIOR APPLICATION NUMBER: 60/056,872
12	PRIOR FILING DATE: 1997-05-23	12	PRIOR FILING DATE: 1997-08-22
13	PRIOR APPLICATION NUMBER: 60/047,618	13	PRIOR APPLICATION NUMBER: 60/056,882
14	PRIOR FILING DATE: 1997-05-23	14	PRIOR FILING DATE: 1997-08-22
15	PRIOR APPLICATION NUMBER: 60/047,503	15	PRIOR APPLICATION NUMBER: 60/056,637
16	PRIOR FILING DATE: 1997-05-23	16	PRIOR FILING DATE: 1997-08-22
17	PRIOR APPLICATION NUMBER: 60/047,592	17	PRIOR APPLICATION NUMBER: 60/056,903
18	PRIOR FILING DATE: 1997-05-23	18	PRIOR FILING DATE: 1997-08-22
19	PRIOR APPLICATION NUMBER: 60/047,581	19	PRIOR APPLICATION NUMBER: 60/056,888
20	PRIOR FILING DATE: 1997-05-23	20	PRIOR FILING DATE: 1997-08-22
21	PRIOR APPLICATION NUMBER: 60/047,584	21	PRIOR APPLICATION NUMBER: 60/056,879
22	PRIOR FILING DATE: 1997-05-23	22	PRIOR FILING DATE: 1997-08-22
23	PRIOR APPLICATION NUMBER: 60/047,500	23	PRIOR APPLICATION NUMBER: 60/056,880
24	PRIOR FILING DATE: 1997-05-23	24	PRIOR FILING DATE: 1997-08-22
25	PRIOR APPLICATION NUMBER: 60/047,587	25	PRIOR APPLICATION NUMBER: 60/056,894
26	PRIOR FILING DATE: 1997-05-23	26	PRIOR FILING DATE: 1997-08-22
27	PRIOR APPLICATION NUMBER: 60/047,492	27	PRIOR APPLICATION NUMBER: 60/056,911
28	PRIOR FILING DATE: 1997-05-23	28	PRIOR FILING DATE: 1997-08-22
29	PRIOR APPLICATION NUMBER: 60/047,598	29	PRIOR APPLICATION NUMBER: 60/056,636
30	PRIOR FILING DATE: 1997-05-23	30	PRIOR FILING DATE: 1997-08-22
31	PRIOR APPLICATION NUMBER: 60/047,613	31	PRIOR APPLICATION NUMBER: 60/056,874
32	PRIOR FILING DATE: 1997-05-23	32	PRIOR FILING DATE: 1997-08-22
33	PRIOR APPLICATION NUMBER: 60/047,582	33	PRIOR APPLICATION NUMBER: 60/056,910
34	PRIOR FILING DATE: 1997-05-23	34	PRIOR FILING DATE: 1997-08-22
35	PRIOR APPLICATION NUMBER: 60/047,596	35	PRIOR APPLICATION NUMBER: 60/056,864
36	PRIOR FILING DATE: 1997-05-23	36	PRIOR FILING DATE: 1997-08-22
37	PRIOR APPLICATION NUMBER: 60/047,612	37	PRIOR APPLICATION NUMBER: 60/056,631
38	PRIOR FILING DATE: 1997-05-23	38	PRIOR FILING DATE: 1997-08-22
39	PRIOR APPLICATION NUMBER: 60/047,632	39	PRIOR APPLICATION NUMBER: 60/056,845
40	PRIOR FILING DATE: 1997-05-23	40	PRIOR FILING DATE: 1997-08-22
41	PRIOR APPLICATION NUMBER: 60/047,601	41	PRIOR APPLICATION NUMBER: 60/056,892
42	PRIOR FILING DATE: 1997-05-23	42	PRIOR FILING DATE: 1997-08-22
43	PRIOR APPLICATION NUMBER: 60/043,580	43	PRIOR APPLICATION NUMBER: 60/057,761
44	PRIOR FILING DATE: 1997-04-11	44	PRIOR FILING DATE: 1997-08-22
45	PRIOR APPLICATION NUMBER: 60/043,568	45	PRIOR APPLICATION NUMBER: 60/047,595
46	PRIOR FILING DATE: 1997-04-11	46	PRIOR FILING DATE: 1997-05-23
47	PRIOR APPLICATION NUMBER: 60/043,314	47	PRIOR APPLICATION NUMBER: 60/047,599
48	PRIOR FILING DATE: 1997-04-11	48	PRIOR FILING DATE: 1997-05-23
49	PRIOR APPLICATION NUMBER: 60/043,569	49	PRIOR APPLICATION NUMBER: 60/047,588
50	PRIOR FILING DATE: 1997-04-11	50	PRIOR FILING DATE: 1997-05-23
51	PRIOR APPLICATION NUMBER: 60/043,311	51	PRIOR APPLICATION NUMBER: 60/047,585
52	PRIOR FILING DATE: 1997-04-11	52	PRIOR FILING DATE: 1997-05-23
53	PRIOR APPLICATION NUMBER: 60/043,671	53	PRIOR APPLICATION NUMBER: 60/047,586
54	PRIOR FILING DATE: 1997-04-11	54	PRIOR FILING DATE: 1997-05-23
55	PRIOR APPLICATION NUMBER: 60/043,674	55	PRIOR APPLICATION NUMBER: 60/047,590
56	PRIOR FILING DATE: 1997-04-11	56	PRIOR FILING DATE: 1997-05-23
57	PRIOR APPLICATION NUMBER: 60/043,669	57	PRIOR APPLICATION NUMBER: 60/047,594
58	PRIOR FILING DATE: 1997-04-11	58	PRIOR FILING DATE: 1997-05-23
59	PRIOR APPLICATION NUMBER: 60/043,312	59	PRIOR APPLICATION NUMBER: 60/047,589
60	PRIOR FILING DATE: 1997-04-11	60	PRIOR FILING DATE: 1997-05-23
61	PRIOR APPLICATION NUMBER: 60/043,313	61	PRIOR APPLICATION NUMBER: 60/047,593
62	PRIOR FILING DATE: 1997-04-11	62	PRIOR FILING DATE: 1997-05-23
63	PRIOR APPLICATION NUMBER: 60/043,672	63	PRIOR APPLICATION NUMBER: 60/047,614
64	PRIOR FILING DATE: 1997-04-11	64	PRIOR FILING DATE: 1997-05-23
65	PRIOR APPLICATION NUMBER: 60/043,315	65	PRIOR APPLICATION NUMBER: 60/043,578
66	PRIOR FILING DATE: 1997-04-11	66	PRIOR FILING DATE: 1997-04-11
67	PRIOR APPLICATION NUMBER		

```

PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,876
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,881
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,909
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,875
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,862
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,887
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/056,908
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/048,964
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/057,650
PRIOR FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: 60/056,884
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: 60/057,669
PRIOR FILING DATE: 1997-09-05

```

Query Match 6.9%; Score 220; DB 10; Length 339;

Best Local Similarity 22.0%; Pred. No. 2.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

38 GEGRNPEVQMSQVLVPCHTNNGELSVGQLKMDITACISARHAG-----CPCVT 91
3 GPDVETPSAIGICRIMRPDANVAGNVHGTLKMEAGALISTRHNSONGECVVAL 62
92 ASMDIYEHTISVQGVNIRAKVRAFNSMEVGIQVASEDLCEKQNNVCALATFYA 151
63 ARVENTDPLSPMCIGEVAVHSAEITYTSKSHVEQVNVNSENILGAKKLTNKATLWYVP 122
152 REITKVKLKQITP---RTEEE---KMEHSVAERRRRLVADTTIKDLANCAIQD 203
123 LSLKNVDKVLVPPVAVSRXEQEEGRKRYEAKLERMETKWRNGDIVQPVLN----- 175
204 LESRDCSRNVPAAKT---RVESVELVLPNHNHOGNTFGGQIMAMNEVATITASRLCR 259
176 -----PEPNTVSYSSQSLIHVGPSCDTLGFVHGVTMKLMDVAGIVAAKHCK 225
260 AHPITKAIEMFHRGSGVGRDLVLKAIYNNAFKSHMEVGVCEAVRQEAETHRRH-INS 318
226 TIVIVASVDAINFHDKIRKGCVITISGRMTFTSNKSMELIVLDADPVVDSSQKRYAAS 285
319 AFMTFVILDADDQPLLPMIRPQPGDGERRYREASAR 355
286 AFTYVYLSQEGRSLEPVQPLVPEDEKKRFEGKGR 322

```

```

RESULT 13
US-10-164-861-488
; Sequence 488, Application US/10164861
; Publication No. US20030225248A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/10/164,861
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: US/09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; NUMBER OF SEQ ID NOS: 757
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 488
; LENGTH: 339

```

```

TYPE: PRI
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (142)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (330)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (335)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (336)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
NAME/KEY: SITE
LOCATION: (339)
OTHER INFORMATION: Xaa equals stop translation
US-10-164-861-488

```

Query Match 6.9%; Score 220; DB 12; Length 339;
Best Local Similarity 22.0%; Pred. No. 2.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

38 GEGRNPEVQMSQVLVPCHTNNGELSVGQLKMDITACISARHAG-----CPCVT 91
3 GPDVETPSAIGICRIMRPDANVAGNVHGTLKMEAGALISTRHNSONGECVVAL 62
92 ASMDIYEHTISVQGVNIRAKVRAFNSMEVGIQVASEDLCEKQNNVCALATFYA 151
63 ARVENTDPLSPMCIGEVAVHSAEITYTSKSHVEQVNVNSENILGAKKLTNKATLWYVP 122
152 REITKVKLKQITP---RTEEE---KMEHSVAERRRRLVADTTIKDLANCAIQD 203
123 LSLKNVDKVLVPPVAVSRXEQEEGRKRYEAKLERMETKWRNGDIVQPVLN----- 175
204 LESRDCSRNVPAAKT---RVESVELVLPNHNHOGNTFGGQIMAMNEVATITASRLCR 259
176 -----PEPNTVSYSSQSLIHVGPSCDTLGFVHGVTMKLMDVAGIVAAKHCK 225
260 AHPITKAIEMFHRGSGVGRDLVLKAIYNNAFKSHMEVGVCEAVRQEAETHRRH-INS 318
226 TIVIVASVDAINFHDKIRKGCVITISGRMTFTSNKSMELIVLDADPVVDSSQKRYAAS 285
319 AFMTFVILDADDQPLLPMIRPQPGDGERRYREASAR 355
286 AFTYVYLSQEGRSLEPVQPLVPEDEKKRFEGKGR 322

```

```

RESULT 14
US-10-094-749-2367
; Sequence 2367, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGIYAMA, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUTKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU

```

```

; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2967
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2967

```

```

Query Match 6.9%; Score 219; DB 15; Length 370;
Best Local Similarity 22.0%; Pred. No. 3.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```

```

QY 38 GEGYRNPTVEVQMSQVLVPCHTNORGLSVGQLKWIIDTTAQLSERRHAG-----CPQVT 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 35 GPDVETPSAIOICRIMRPDANVAGNVHGGTILKMIAGAKITRHCNSONGECVVAL 94
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 ASHDDIYFHTISVGQVNTKAKVNRAPNSMEVGIQVASEDLCEKQMNVCALATFVA 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 ARVERDTELPSCMGVAHVASEAETTSKHSVEYQVVMSENILTGAKKLTNKTLMVVP 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 RREITTKYKQIOTP-----RTEERKMEHSVAERRRMLVY--ADITKOLLANCAIQGD 203
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 LSLKNVDKVLVEPVVSVROEQEEGRKRYEAQKLEMETKMRNGDVIQPVLN----- 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 LESRDCSRMVAEKT-----RVESVELVLPRIHNGQNTFGSQIMAMENVATTIASRLCR 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 208 -----PEPTVSYSSQSLIHVGFSDCTIHGVHGGVTKMLDEVAQIVAAHCK 257
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 AHPTLKAIEHFHFGPSQVDRVLKAIYNNAPKSHMEVGVCEVAEQEAETRRH--INS 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 258 TNITVTSVDAINFHDKIRKGCVITITSGRMFTFSNKSKEIVLVADVPVDSQKRYTAA 317
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 319 AFMTFVLVLDADDQQLPWIRPQPGDGRERYRARSAR 335
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 318 APTTVSLSQEGRSLVPQVLPETEDKKRFBEGKGR 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

US-10-282-122A-55313
; Sequence 55313, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55313
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-282-122A-55313

```

```

Query Match 6.5%; Score 207.5; DB 12; Length 159;
Best Local Similarity 33.3%; Pred. No. 1.2e-12;
Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;

```

```

QY 226 VLPRIHNGQNTFGGQIMAMENVATTIASRLCRAPHTLKAIEHFHFGPSQVDRVLK 285
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 24 IFPDINTNNTITGGILMSILDRILAVARRHCESICVTALVDVRFYAPAYMGENILCC 83
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 AIYNNAPKSHMEVGVCEA---YRQEAETRRHNSAFMTFFVLVLDADDQQLPWIRPQ 342
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 84 ASVNRSMRISLSEGVKMAENITKQ---QRHTSAVFTFVAVDENNSPVEIPELVET 139
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 GDEGRYRARSARKIRL 360
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 QSEIRRFHEADQRRARL 157
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: August 12, 2004, 15:51:19
Job time : 50 secs

Blank Sheet

Db 1 MFSNRTSRKSIHPESGDDPTMAEGEGYRNPTVEQMSQVLVPCHTNHRGELSIGQLKWI 60
 QY 74 DTTAGLSAERHACPCVTASMDIYFHEHTISVGQVNIKAKVRAFNSSMEVGIOVASED 133
 Db 61 DTTAGLSAERHACPCVTASMDIYFHEHTISVGQVNIKAKVRAFNSSMEVGIOVASED 120
 QY 134 LCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD 193
 Db 121 LCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD 180
 QY 194 LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIA 253
 Db 181 LTHTCVQDDLD-KDCSNMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIA 239
 QY 254 ASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR 313
 Db 240 ASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR 299
 QY 314 RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV 373
 Db 300 RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV 359
 QY 374 PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVA 433
 Db 360 ALSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVA 418
 QY 434 VDAQAFLSLDLRQRPBMDKHYSVELVQVDEDDAIYHTSPALGHTKPODFVILLAS 493
 Db 419 VDAQAFLSLDLRQRPBMDKHYSVELVQVDEDDAIYHTSPALGHTKPODFVILLAS 478
 QY 494 RRKPCDNDPPYIALRSTVLPHTREPYRREGELTCSGCLMREBDQTLKCCWVVSJTE 553
 Db 479 RRKPCDNDPPYIALRSTVLPHTREPYRREGELTCSGCLMREBDQTLKCCWVVSJTE 533
 QY 554 LVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR 600
 Db 534 -QATPGFLNY-VTTNVSGLSSEFYN-----TFKACESFLDNR 569

RESULT 2
 ID 080TT9 PRELIMINARY; PRT; 561 AA.
 AC Q80TT9;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MKIAA0707 protein (Fragment).
 GN MKIAA0707.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Imanoto S., Aizawa H., Yuasa S.,
 Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologues
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:35-48(2003).
 RL EMBL; AK122350; BAC65632.1; -;
 DR GO:0003824; F: catalytic activity; IEA.
 DR InterPro: IPR002913; START.
 DR InterPro: IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR SMART; SM00234; START; 1.
 DR PROSITE; PSS0948; START; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 561 AA; 63654 MW; 6D7183C48F103EAA CRC64;

Query Match 76.1%; Score 2421; DB 11; Length 561;
 Best Local Similarity 82.3%; Pred. No. 1.3e-201;
 Matches 469; Conservative 33; Mismatches 47; Indels 18; Gaps 5;

QY 35 MADGEGYRNPTVEQMSQVLVPCHTNHRGELSIGQLKWI DTTAGLSAERHACPCVTASMDIYFHEHTISVGQVNIKAKVRAFNSSMEVGIOVASEDLCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 4 MAEGGYRNPTVEQMSQVLVPCHTNHRGELSIGQLKWI DTTAGLSAERHACPCVTASMDIYFHEHTISVGQVNIKAKVRAFNSSMEVGIOVASEDLCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 95 DDIFYEHTISVGQVNIKAKVRAFNSSMEVGIOVASEDLCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 64 DDIFYEHTISVGQVNIKAKVRAFNSSMEVGIOVASEDLCSEKQNVCKALATFVAREITVVKIKQITPRTEEEKMHSVAERRMRLVADTIKD LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 155 ITTKVLCQITPRTEEEKMHSVAERRMRLVADTIKD LLMNCAIQGLLESRCGRMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 124 LSKYKLMQVLPTEEEKTEHGAERRMRLVADTIKD LTHTCVQDDLD-KDCSNMYPAEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 215 AEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 183 AEKTRVESVELVLPphanQGNTFGQIMAMMENVAITIAASRLGRAHPTLKAITEMHFRGSGVQDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 275 PSQVGDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 243 PSQVGDRLVKAIVNNAFKHSMVEGVCVEAYROEATHR RHINSAMTFVVLADDPQPLPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 303 LPWIRPQDGERRYEASARKKIRLDRKYIVSCQTEV PLSPWMDPSNOVYLSYNNVSSLKMLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 395 LKMTLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 363 LKMTLVAKDNWVLSSEISQVRLYTLDEDKFLSFHMEVVALVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 Db 455 HYSVELVQVDEDDAIYHTSPALGHTKPODFVILLASRRKPCDNDPPYIALRSTVLPHTREPYRREGELTCSGCLMREBDQTLKCCWVVSJTE
 QY 422 HYSVELVQVDEDDAIYHTSPALGHTKPODFVILLASRRKPCDNDPPYIALRSTVLPHTREPYRREGELTCSGCLMREBDQTLKCCWVVSJTE
 Db 515 THRETPYRREGELTCSGCLMREBDQTLKCCWVVSJTE LVSAGFYSGWGLSESRKGRSDGMNGKLAGHSLTKAIPVAKINSR
 QY 482 THRETPYRREGELTCSGCLMREBDQTLKCCWVVSJTE -QATPGFLNY-VTTNVSGLSSEFYN-----TFKACESFLDNR 551
 Db 575 DGMNGKLAGHSLTKAIPVAKINSR 600
 QY 535 EFYN-----TFKACESFLDNR 551

RESULT 3
 ID 08NM17 PRELIMINARY; PRT; 339 AA.
 AC 08NM17;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase.
 GN CGL2586.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OC NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF005282; BAB99979.1; -;
 DR GO:0016787; F: hydrolase activity; IEA.
 DR InterPro: IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR Hydrolase; Complete proteome.
 KW SEQUENCE 339 AA; 37521 MW; A2305D953D604AFD CRC64;

Query Match 8.0%; Score 255; DB 16; Length 339;

Best Local Similarity 25.7%; Pred. No. 1.6e-13;
Matches 89; Conservative 56; Mismatches 171; Indels 30; Gaps 10;

QY 41 YRNPEYQMSQVLVPCHTNQGELSV--GOLIKWIDTTACLSAEHACPCVTASMDITY 98
Db 14 HRSP-EVVLRFMAAEDTDLVLMAGSHGVGGRLVEMIDKAAVACATQMSGYCTTAAVGHII 72
QY 99 FEHTISVGVVNIKAKVARNFSSMEVGIQVASELCEKQNVKALATVAREIT-- 156
Db 73 FTRPISGHVVEVSRIMTGRSSMHIVNLVSLADPRDQNTYRACDCIVIVAKTATGR 132
QY 157 KVCLKQITPRTTEEEKMEHVAERRRRLVYADTIKDLANCAIQGDESDCSMVAAE 216
Db 133 ATPVPSFTPKNEBEGRLVLEANSRIGLRK-----ALAEWEKQTYNG--PSE 177
QY 217 KTRVESVELVLPPIAHNQGNTFGQIVAMMENVAITIASRLCRAHPTLKATIEHFRRGPS 276
Db 178 APRILTRFLAKPTDINMGKVGHTAMEMIDEAGAACTMWSGNHTVAVAGIRFPQPI 237
QY 277 QVGRRLVKAIVNNAFHSMEVGV--EAYRQEAETHRR-HINSAPMTFVLDADDQ 332
Db 238 QIGLIEVDARMMFTDKRSQWMSIHVAGDAHRGRABLETAIH--ATVYLGIDVDGEP 294
QY 333 QLLPWIRPQDGERRRYREASARKKIRLDRKYI-VSCQTEVPLSV 377
Db 295 LPAQFVPRPREDIQLEHANIILDLAD--YTPMPLFORRVPLQI 338

RESULT 4

Q8FMM5 PRELIMINARY; PRT; 335 AA.
AC O8FMM5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN CE2478.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawaiyaei Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.,
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005222; BAC19288.1;
DR GO; GO:000324; Catalytic activity; IEA.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 335 AA; 36744 MW; 6985934F1F837D76 CRC64;

Query Match 7.3%; Score 233; DB 16; Length 335;
Best Local Similarity 23.7%; Pred. No. 1.2e-11;
Matches 79; Conservative 55; Mismatches 178; Indels 22; Gaps 6;

QY 35 MADGEGYRNP--TEYQMSQVLVPCHTNQGELSV--GOLIKWIDTTACLSAEHACPCPV 90
Db 1 MSDNQTASPRSPDITLRLAAPTDLVLMAGSHGVGGRLVEMIDKAAVACATQMSGYCTV 60
QY 91 TASMDIYFEHTISVGVVNIKAKVARNFSSMEVGIQVASELCEKQNVKALATVAREIT 150
Db 61 TAYVGHIFHTRPIPSGHVVEVSRIMTGRSSMHIVNLVSLADPRDGTTRACDCIVITV 120
QY 151 ARRET--KVCLKQITPRTTEEEKMEHVAERRRRLVYADTIKDLANCAIQGDESND 208
Db 121 AADYATGKTQEVPTFVPRKTAEEERVLQSLSRIDLR-----ALAEWEKQI 167
QY 209 CSRWVPAEKTRVESVELVLPPIAHNQGNTFGQIVAMMENVAITIASRLCRAHPTLKALE 268

Db 168 YDG--PSDAPRLITRFLAKPTDINMGKVGHTAMEMIDEAGAACTMWSGNHTVAVAG 225

QY 269 MEHRRGSGVQDRLVKAIVNNAFHSMEVGVCEVAYKQE-AETHRRHINSAEFTVLD 327

Db 226 GIRFNPISTIDLEVDRLMRTRDTRSQWMSIHVRSNGPKGRDHLRTAIAHTVYIGID 285

QY 328 ADDQPOLPWIRPQDGERRRYREASARKKIRLD 361
Db 286 LDGQPLAPPTPTPTPEDIQLEHANIILDLRAE 319

RESULT 5

Q97B23 PRELIMINARY; PRT; 165 AA.
AC Q97B23;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Acyl coenzyme A thioester hydrolase.
GN TV0636 OR TVG0629193.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatiales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima T., Watanabe K., Yamazaki M., Kanohori K., Kawamoto T.,
RA Numashita T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AF000993; BAB59778.1;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 165 AA; 19195 MW; 5392888A47B2768D CRC64;

Query Match 7.2%; Score 228; DB 17; Length 165;
Best Local Similarity 34.6%; Pred. No. 1.2e-11;
Matches 55; Conservative 35; Mismatches 61; Indels 8; Gaps 5;

QY 214 PAE-KTRVESVELVLPPIAHNQGNTFGQIVAMMENVAITIASRLCRAHPTLKAIEMFHF 272
Db 6 PSESETFVE--RMVLPEMDIDYEVLYGGRLMEWIDNCASIVATKCKRKRIVTGSIDSLF 63

QY 273 RGPQGVQDRLVKAIVNNAFHSMEVGVCEVAYKQAEFH-RRHINSAPMTFVLDADDQ 331
Db 64 LPIHIGDWVLLHGTYNTTKSTMEI--IDVIKQGLGIRYATKAYLIVADADGR 121

QY 332 POLPWIRPQDGERRRYREASARKKIRLDRKYVSCQ 370
Db 122 PTEIPQIEPTDDEKRYVAEAKRAER--RKLEITIKQ 158

RESULT 6

Q8U9A2 PRELIMINARY; PRT; 276 AA.
AC Q8U9A2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Acyl-CoA hydrolase.
GN ATU3826 OR AGR 1.2016.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boye D., Grant C.,
 RA Chapman P., Clendenning U., Deatherage G., Gillet W., St. A.,
 RA Kulyavin T., Levy R., Li M.-U., McCelland E., Palmieri A., Gordon D.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Perry M.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepano W., Bery M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.,
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58.";
 RL Science 294:2317-2323 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu P.,
 RA Follam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328 (2001).
 DR EMBL; AF009314; AAL44636.1; ALT_INIT.
 DR EMBL; AF008300; AAK89583.1; -.
 DR PIR; AF3027; AF3027.
 DR PIR; E98257; E98257.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 KW Hydrolase; Complete proteome.
 SEQUENCE 276 AA; 29993 MW; 42031EB95951E76E CRC64;

Query Match 7.1%; Score 224.5; DB 16; Length 276;
 Best Local Similarity 25.2%; Pred. No. 5.1e-11;
 Matches 72; Conservative 42; Mismatches 127; Indels 45; Gaps 5;

QY 48 QMSQVLCHTNQRELSGQLKWIPTTACLSARHNGCCPVTSAMDYIEHTISVQ 107
 DB 18 RLIDIVFPGDTHHGGTIGGTGLAMDRVAFAIRFGPTPTVTSACRIDRQARICH 77
 QY 108 VVNIKAVYARFNSMEVGIQVASEDLCEKQNVCKALATFVARREITKVKLKITPRT 167
 DB 78 IVEFTARPKAKRRRLTYVEVWVAETIIGRQHTCRGIFHVA-----IPRG 125
 QY 168 EEEKMEHSVAAERRMRLVYADTICDLANCAIQDLESRDCSRWPAKTEVESVELYL 227
 DB 126 ED-----AASYVLPBELTET-----PDPSDAVTWEIVF 155
 QY 228 PEHANHGNTFGQGMAMMENVAATIASRLCRAPHTLKAIMEHFPGQVGDRLTKAI 287
 DB 156 PQANASAGMEGEALAYTKAFYAAASYCKGLVILASERKIDPRALEIGIYEAQKH 215
 QY 288 VVNAFGHMEVGVCEAVYRQEAETRRHIN-SAFMTFVVLADDDQ 332
 DB 216 VERRVSSMSIQ-TKLWSENLLTGRHITATGFTMVAVDDKHRR 259

RESULT 7

OS 09HKJ9 PRELIMINARY; PRT; 165 AA.
 AC 09HKJ9;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase related protein.
 GN TA0599.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasma; Thermoplasmatales;

OC Thermoplasmataceae; Thermoplasma.
 CX NCBI_TaxID=2303;
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513 (2000).
 DR EMBL; AL445064; CAC11738.1; -.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hydrolase; Complete proteome.
 SEQUENCE 165 AA; 19233 MW; 28F1614521FPA4DA CRC64;

Query Match 6.8%; Score 217.5; DB 17; Length 165;
 Best Local Similarity 32.3%; Pred. No. 9.6e-11;
 Matches 50; Conservative 38; Mismatches 60; Indels 7; Gaps 4;

QY 217 KTRVSEVELVPPAHNGQNTFGQGMAMMENVAATIASRLCRAPHTLKAIMEHFPGPS 276
 DB 10 ETVYE--RMVLPEDMDIYELYGRLEMIDNCASIVATKCRKRTVGSIDSLFFLIPI 67
 QY 277 QVGRIVLKAIVNAFKHMEVGVCEAVYRQEAETH--RRHNSAFMTFVVLADDDQPOL 335
 DB 63 HLGDWVILHGYINTYTKSTWTE--IDVKEGLTGIRYATKALYTVAVLSDGRPREI 125
 QY 336 FWIRPQDGERRYEASARKKIRIDRYIVSCQ 370
 DB 126 PQIVETDEKRRYQDAEKRAER--RRLESIKQ 158

RESULT 8

OS 09KEQ1 PRELIMINARY; PRT; 157 AA.
 AC 09KEQ1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Acyl-CoA hydrolase.
 GN BH0798.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=86665;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331 (2000).
 DR EMBL; AR001509; BAB04517.1; -.
 DR PIR; F83749; F83749.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hydrolase; Complete proteome.
 SEQUENCE 157 AA; 17328 MW; 8639B00827EEBA3A CRC64;

Query Match 6.8%; Score 216.5; DB 16; Length 157;
 Best Local Similarity 32.0%; Pred. No. 1.1e-10;
 Matches 48; Conservative 32; Mismatches 67; Indels 3; Gaps 2;

QY 214 PAEKTVESVELVLPANHGQNTFGQGMAMMENVAATIASRLCRAPHTLKAIMEHFPR 273
 DB 6 PYRRSRTIGRLVLPDNLHGLTIGSKVLAIDEIATLAKMHANAAVYTAISIDVDFK 65

Db 73 LASVNTYNTSMRGKIVVTENIRERSV---RHNSCFVTWAVDNRKPAVPEPEPT 129
 Qy 343 GDGRRRYREASARKIR-LDRKY 364
 Db 130 PEGRKRRTQAGRRQRIQOELEKRY 153

RESULT 12

Q92TG6 PRELIMINARY; PRT; 167 AA.

AC Q92TG6; (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Putative acyl-CoA thioester hydrolase protein (EC 3.1.2.-).
 GN R81554 OR S821634.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymB (megaplasmid 2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 [1]
 RN SEQUENCE FROM N.A.

RC STRAIN=1021;
 RC MEDLINE=21396508; PubMed=11491431;
 RA Vishan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
 DR EMBL; AL603647; CAC94953.1; -.
 DR PIR; A96036; A96036.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr. supf.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hydrolyase; Plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 167 AA; 18806 MW; B785C8DBE0A0F0C CRC64;

Query Match 6.4%; Score 204; DB 16; Length 167;
 Best Local Similarity 32.2%; Pred. No. 1.5e-09;
 Matches 48; Conservative 31; Mismatches 68; Indels 2; Gaps 2;

Qy 212 WPAKTR-VSEVELPPEPHANHGNTFGQIMAMENAVTTASRLCRAPHTLKAIEWF 270
 Db 1 MMTBETHTLVMTVMTPTPMANFGKVGGLALINDRYAFSCASRFKQYAVTLSDQV 60
 Qy 271 HFRGPGVGDRLVTKAIVNNAFKHMEVGVCEAYRQGEATHRHNSAFMTFVLDADD 330
 Db 61 IFKEPIHVGELVTFPAALIVAGRTSMEVGRVBAENIRAGT-RRHTNSCYFTWAVDDAG 119
 Qy 331 QPQLPWIRPQDGGRRRYREASARKIR 359
 Db 120 RPTVPEYHPTATEKRRHQAAELKRTLR 148

RESULT 13

Q97WD9 PRELIMINARY; PRT; 318 AA.

AC Q97WD9; (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Probable acyl-coenzyme A thioester hydrolase.
 GN S802287.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chen-Weiner C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Eranuo G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.B., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).
 DR EMBL; AE006832; AAK42449.1; -.
 DR PIR; B90399; B90399.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr. supf.
 DR Pfam; PF03061; 4HBT; 2.
 KW Hydrolyase; Complete proteome.
 SQ SEQUENCE 318 AA; 36244 MW; 97D52EPD4CD33AE4 CRC64;

Query Match 6.4%; Score 204; DB 17; Length 318;
 Best Local Similarity 22.0%; Pred. No. 3.8e-09;
 Matches 72; Conservative 67; Mismatches 132; Indels 56; Gaps 9;

Qy 58 TNORGELSVGQLKMDITTAQLSARRHAGCPCTVTSMDITFEHTISVGQVNNIKAKNR 117
 Db 22 SNFLRLHGGDMLEFVETGMISAKVAMGTTLASLDVVEKPKVRIGDITKVRKATVY 81
 Qy 118 AFNSMVEIGIOVABEDLCEKQNMVCKALATFVARREI---TKVTKQITPTEEEKMEH 174
 Db 82 IGNSTLEVEYAVFVRD-----EEVSAVATVYKVDLLRTPPVVK-IMAESEDDKKI 134
 Qy 175 SVAERRRMRV-----VYADTIKDLLANCAIQGLDSRDCSRWVPAKTRVESVE 224
 Db 135 DEAKKRNRRLSKTYNRQKMFYVDIDTGL-----RYRISNV- 172
 Qy 225 LVLPPEPHANHGNTF-GGQIMAMENAVTTASRLCRAPHT-----LKAIEMHF 272
 Db 173 IHVSPELTYDGRISAGKILKIMDLGIIICIRYLDHNSNLYDNSFNAAVTVAAKGLAF 232
 Qy 273 RGPSPGVGRVLVTKAIVNNAFKHMEVGVCEAYRQGEATHRHNSAFMTFVLDADDP 332
 Db 233 YSPRLNIIILRA--GLVYGNISADIIINVRBEDLNGAKHVAATFTYRVKDGKRP 290
 Qy 333 QLPWIRPQDGGRRRYREASARKIR 359
 Db 291 IKIPRYTPVTRERKRLYEALTRRGLR 317

RESULT 14

Q81RJ1 PRELIMINARY; PRT; 168 AA.

AC Q81RJ1; (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Cytosolic long-chain acyl-CoA thioester hydrolase family protein.
 GN BA2053.
 OS Bacillus anthracis (strain Ames).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=198094;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22608414; PubMed=12721629;
 RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
 RA Nelson K.E., Tettelin H., Routs D.E., Eisen J.A., Gill S.R.,
 RA Holtzapfel E.K., Okstad O.A., Helgason E., Ralston J., Wu M.,
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radeue D.,
 RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Niernan W.C.,
 RA Hazen A., Cline R., Redmond C., Tiwari J.E., White O., Salizco S.L.,
 RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolisto A.-B.,

RA Fraser C.M.;
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to
 RT closely related bacteria."
 RL Nature 423:81-86(2003).
 DR EMBL; AE017030; AAP25941.1; -.
 DR TIGR; BA2053; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 1.
 KM Hydrolase; Complete proteome.
 SQ SEQUENCE 168 AA; 18926 MW; 87478FDE8C4ACB34 CRC64;

Query Match 6.4%; Score 203; DB 16; Length 168;
 Best Local Similarity 35.0%; Pred. No. 1.8e-09;

Matches 55; Conservative 27; Mismatches 71; Indels 4; Gaps 2;

QY 43 NPTVOMSQLVLPCHTNGELISVGLKMWIDTACLSAERHAGCPCTASMDIYEHT 102
 DB 10 NESRVFKTSRVFPDINDHNTLPFGKILAEWDVYASISATRHAKRECVTASMDWDFLHP 69

QY 103 ISVQGVNIRAKVRAAFNSMEVGIOVASEDLCSKQNNVCKALATFVA-FREITKYKX 161
 DB 70 VRSSDCVSYSEFVMTGRTSMVEVFKVAEDLISGEKRIATSFVTFVALSKENNPVPVP 129

QY 162 QITPRTEEEKMEHSA--AERRRRLVYADTIKDL 195
 DB 130 RVIPETBEKEKLRIVAVLRABQRHTRAKESKRVATLL 166

RESULT 15

Q9HT54

ID Q9HT54 PRELIMINARY; PRT; 188 AA.

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein PA5519.

OS *Pseudomonas aeruginosa*.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20457337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Keizer J., Sailer M.H., Hancock R.E.W., Iori S., Olson M.V.;

RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964 (2000).

DR EMBL; AE004964; AAC08904.1; -.
 DR PIR; F82956; F82956.

DR GO; GO:0003824; F:catalytic activity; IEA.

DR InterPro; IPR006683; Thioestr_supf.

DR Pfam; PF03061; 4HBT; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 188 AA; 21215 MW; 8705C22F1C9F847E CRC64;

Query Match 6.3%; Score 201; DB 16; Length 188;

Best Local Similarity 30.2%; Pred. No. 3.2e-09;

Matches 54; Conservative 36; Mismatches 69; Indels 20; Gaps 3;

QY 13 LASVSNRKTS-----RKSLRAANDSAMADGGRNPTVQMSQIVLPCHTNGELISV 66

DB 9 LPSSPASVSVSPARPERKASAMQPGNN-----QLSMIVLMTDPDMANFSGNVHG 55

QY 67 GQLTKMIDTTACTLSAERHAGCPCTASMDIYEHTISVQGVNIRAKVRAAFNSMEVG 126

DB 56 GTLLKYIDVAVACASRYAGRVVTLASDVQVIFREPVHVGELVTFELASVNTGRTSMELG 115
 QY 127 IOVASEDLCSKQNNVCKALATFVARREITK-VLKQITPRTEEEKMEHSAERRRMR 184
 DB 116 VKVITENIREQSVHHTNSCFFTWALDDEKRPVAVRPLETETABQKRRFAOQGRQR 174

Search completed: August 12, 2004, 15:45:21
 Job time : 44 secs

Blank sheet

; EARLIER FILING DATE: 1997-03-07

RESULT 5
US-09-134-001C-4696
; Sequence 4696, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4696
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4696

Query Match 5.9%; Score 186.5; DB 4; Length 177;
Best Local Similarity 30.6%; Pred. No. 1.3e-12;
Matches 56; Conservative 30; Mismatches 78; Indels 19; Gaps 4;

QY 17 FSNRTSRKSLRAAGNDAMADGEGYRNFTEVOMQVLPCHTNOGELSVGQLKMDIT 76
DB 1 FMAINTNRS-----KKMSDAKCIK-----TROVPODTHHTTFFGTLMANIDEL 46
QY 77 ACLSAERHAGCPCTASMDIYFEHTISVGQVNNKAVNAPFNSMVGIOVASEDICS 136
DB 47 AAITAMKRAQNPVASTDSVDFPLPITGDLSTEAANVSTAGTSSMTCVOIVDDYK 106
QY 137 EKONVVCALATFYAREITK-VKLQITPREEKMEHSA---AERRRRLVYADTI 191
DB 107 NERHIAALSFITFVALDSDGKPTSPVDVEPETSVEKMEHAPQVARKERKESIDTI 166
QY 192 KDL 194
DB 167 EYL 169

RESULT 6
US-09-252-991A-31606
; Sequence 31606, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31606
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31606

Query Match 5.7%; Score 180.5; DB 4; Length 167;
Best Local Similarity 33.6%; Pred. No. 5.7e-12;
Matches 47; Conservative 25; Mismatches 67; Indels 1; Gaps 1;
QY 46 EVOMQVLPCHTNOGELSVGQLKMDITTAACLSAERHAGCPCTASMDIYFEHTISV 105
DB 8 QLTMSVLTMPDRANFGVNHGALLKLDEVAFAACKRYAGRVVLTSLVDVYFRPDIHV 67

QY 106 GOVNNIKAVNAPFNSMEVGIQVASEDLCSKQMNVCALATFYAREITK-VKLQIT 164
DB 68 GELVTFASVNYGRTIMEVGAVKMTNHERVYRHINSCEFTIVVAMDORPVAVPLE 127
QY 165 PREEEMHESVAERRMR 184
DB 128 PETAIEKRYAQAALAREQR 147

RESULT 7
US-09-540-236-2636
; Sequence 2636, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2636
; LENGTH: 162
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2636

Query Match 4.5%; Score 143; DB 4; Length 162;
Best Local Similarity 38.1%; Pred. No. 1e-07;
Matches 32; Conservative 17; Mismatches 35; Indels 0; Gaps 0;

QY 45 TEVOMQVLPCHTNOGELSVGQLKMDITTAACLSAERHAGCPCTASMDIYFEHTIS 104
DB 60 TDLNITVMTDMANFSGNHGTLIKLDDVAAACASRYAGVYVLTSLVDVKEPELY 119
QY 105 VGQVNNIKAVNAPFNSMEVGIQ 128
DB 120 VGEIVTFIARINHVGTSMEVGIR 143

RESULT 8
US-09-543-681A-4296
; Sequence 4296, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4296
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4296

Query Match 4.2%; Score 135; DB 4; Length 143;
Best Local Similarity 31.2%; Pred. No. 6.7e-07;
Matches 39; Conservative 14; Mismatches 62; Indels 10; Gaps 3;

QY 218 TRVESVELVLPNHNQNTFGQIMAMNENVAITIASRLCAHPTLKAIEHFHFGSQ 277
DB 22 TRVSKV--LPTTNHSTLFGDTALAMWDVSPITATPCKRKLVTUSTEKINFHPLP 79
QY 278 VGBRLVTKAVNNAPFGHMEVGVCVAVROEATRRRHINSAFMEFVLADAD-----D 330
DB 80 SGTIIILVGEVIRVGTSLTVNVSIFLEMYVAGREBVHGGF--NVVAADHGKPTPLFD 138
QY 331 QPQLT 335

DB 139 KPSIL 143

RESULT 9
US-09-328-352-5165
Sequence 5165, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5165
LENGTH: 145
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5165

Query Match 3.8%; Score 121; DB 4; Length 145;
Best Local Similarity 23.6%; Pred. No. 2.7e-05;
Matches 35; Conservative 32; Mismatches 63; Indels 18; Gaps 3;

QY 39 EGRNPTEV-QMSQVLPCHTNORGEISVQQLKMTDTTACLSAERHAGPCVTASMDDI 97
DB 8 EATYKPEGLISLTQIAMPADTNWSGDVFGGIWISQMDLAGAIIHERFSRGRCATISINQM 67

QY 98 YFHTISVGVVNIKAKVNAFNSMEVGIQVASEDCSRKQNVCAALTFVA----- 151
DB 68 TELVPKGVGVISCTYKILNVGNSTISQMLEVDSDHSRAPRIITGEVTFVAVDYKGN 127

QY 152 RREITVKLQKQITPREEEMHESVAE 179
DB 128 KRQI-----PEAKQKFLISGE 144

RESULT 10
US-08-691-814B-6
Sequence 6, Application US/08691814B
Patent No. 5981218
GENERAL INFORMATION:
APPLICANT: Rio, Marie-Christine
APPLICANT: Tomasetto, Catherine
APPLICANT: Baasset, Paul
APPLICANT: Byrne, Jennifer
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful
TITLE OF INVENTION: as Leukemia Markers and in Breast Cancer Prognosis
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave, NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691.814B
FILING DATE: 31-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,183
FILING DATE: 09-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1383.0090001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2543
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-691-814B-6

Query Match 3.7%; Score 118; DB 2; Length 445;
Best Local Similarity 20.0%; Pred. No. 0.00047;
Matches 70; Conservative 53; Mismatches 97; Indels 130; Gaps 17;

QY 282 LVKAIYN-----NAFHSSEVGVCEVAYRQEAETRRHINSAPMTFVILDADQPOLLP 336
DB 135 LIVKVIISLILSKAFGYLLPIVSVFLAW-----LETWFLDFKVL----- 174

QY 337 WIRPQDGGRRY-----REASARKI 358
DB 175 ----PQAEERRWYLAQVAVARGLPFGALISGQFYSPESFAGSDNESDEVAAGKSF 231

QY 359 RL-DRXYIVSCKQTEVPLSVWDPNSQVYLSYNNVSLAKMLVAKDNWVL--SEISQVRL 415
DB 232 SAQREXYIROQEKATAVD-----QLAQEENKFKENNEYGDT-V 271

QY 416 YTLIEDKFLSFNHEMVHV-D-AAQATLLSDLRQPE-----NDKYSVELVQVDEDD 469
DB 272 YTLR---VPFHGKTFILKTFELPCPAELVYQEVTLQPERWVLMNKVTACQIIQRVBDNT 327

QY 470 AI-YHTSPALGGHTKPODFVIL--ASRRRCQNGDPY-----VIALSVTLPTRETPY 522
DB 328 LISVDSAGAGGVNSPRDQVNRITERRR---DRYLSGATSTSHAKPPTH-----Y 378

QY 523 RRGETLCSGFCILWRBGDLTKCQVRSVLTSLVTSAGSGFYSWGLESRSKR 572
DB 379 VRGENGRGFTIVLKSASNPVCTFV-----WILNTDLKGR 413

RESULT 11
US-09-252-991A-18251
Sequence 18251, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18251
LENGTH: 312
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18251

Query Match 3.6%; Score 113; DB 4; Length 312;
Best Local Similarity 33.3%; Pred. No. 0.00092;
Matches 26; Conservative 8; Mismatches 36; Indels 8; Gaps 1;

QY 212 MVAPEKTRVESVEL-----VLPRHANQGTGCGQIMAMNENATTAASPLCAHPT 263
DB 83 MTPREGEILARTELSTRTKAVFPPTNNHNTLROGTALAMWDEVSTIATRFCLPLV 142
QY 264 LKAIEMFHRGPOVQDR 281

Db 143 TVSTDRIDFNHPICGAR 160

RESULT 12

US-09-543-681A-4796
; Sequence 4796, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4796
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4796

Query Match 3.5%; Score 112.5; DB 4; Length 146;

Best Local Similarity 26.5%; Pred. No. 0.00026;

Matches 36; Conservative 30; Mismatches 61; Indels 9; Gaps 2;

QY 46 EVQMSQVLPCHTNORGELSVGQILKIDTTACTSABRHAGCPCVTASMDIYFEHTISV 105

Db 14 ELVIRTLAMPADTNANGDIFGGMWMSQMDIGAILAKEIALGRVTVSVTGITFLKPPAV 73

QY 106 GGVNNIRAKVNRANSSMEVGIOVASDLCSE---KQNVCKALATFPAAREITFVKLKQ 162

Db 74 GDVVCVYARCKLTGSSITVINEVWVKVATEPVQGRYRATFAYVTVAVDENNSAR--- 130

QY 163 ITPRTEERMEHVSVA 178

Db 131 ---RLPBGKTNFTLAS 143

RESULT 13

US-09-489-039A-7629
; Sequence 7629, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7629
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7629

Query Match 3.4%; Score 109.5; DB 4; Length 133;

Best Local Similarity 27.5%; Pred. No. 0.00048;

Matches 30; Conservative 26; Mismatches 50; Indels 3; Gaps 1;

QY 46 EVQMSQVLPCHTNORGELSVGQILKIDTTACTSABRHAGCPCVTASMDIYFEHTISV 105

Db 12 EMVIRTLAMPADTNANGDIFGGMWMSQMDIGAILAKEIALGRVTVSVTGITFLKPPAV 71

QY 106 GGVNNIRAKVNRANSSMEVGIOVASDLCSE---KQNVCKALATFVA 151

Db 72 GDVVCVYARCKLTGSSITVINEVWVKVATEPVQGRYRATFAYVTVAVDENNSAR 120

RESULT 14

US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: NO. 5994070el TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHAYE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,267
; FILING DATE: 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,214
; FILING DATE: 27 MARCH (1996)
; ATORNEY/AGENT INFORMATION:
; NAME: Amy B. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2860 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-267-2

Query Match 3.3%; Score 106.5; DB 2; Length 2860;

Best Local Similarity 19.8%; Pred. No. 0.3;

Matches 126; Conservative 89; Mismatches 227; Indels 193; Gaps 32;

QY 22 SRKSLRAGNDSAMADGEGYRNP-----TVQMSQVLPCHTNORGELSVGQILK 71

Db 1087 SAKQLEWHID---NGSEFYLSHTSTGSSIOHTOELKEHEFEQITAKQTERYKLIQ 1142

QY 72 WIDTACTSABRHAG--CPCVTASMDIYFEHTISVGQVNNIRAKVNRANSSMEVGIOV 129

Db 1143 LAD-GFCCKGHAHAELIKCVTA-VDKRYRPSLSM-----EKYRSTLEKALGI 1189

QY 130 ASEDLCSEKQNVCKALATFPAAREITFVKLKQITPRTEERMEHVSVAERRMALVYD 189

Db 1190 SSDSNKSSKSLQ---LDIIPASIPGSEVKLRDAHAELNEERKSA---RRKEPTIME 1240

QY 190 TKDILLACALQGDLESRCSEMVPAEKTREVSVELVPRPHANQCNFTFG--QTMAMWE 247

Db 1241 LIQ---TEKAYVRD--RECDTYLMENT--SGVEIIPGILYINKELLIFGNNOIEIYEFIN 1293

QY 248 NVATTAASRLCRAPHTLKAIEFHRGPSQVGDRLVLKALVNNAPGHSMEVGVC----- 301

Db 1294 NI-----FLKELEKYE-QLPEDVG---HCFVTWADKFCQYVTVYCKRKPD 1334

QY 302 VEAYIOEA-----ETHRRH-INSAPMTFVVDADDOQPOLLPITRQPGGGERYREASAR 355

Db 1335 TQILIEHGSYDEDIQORHGLANSISYDI-----KVVORI 1370

QY 356 KRIPLDRKYIVSCQK-----TEVPLSVPMDPENQVYLSY-----NNVSLKMLVAK 401

Db 1371 TYQULLKELLTCCEGKGEIYDGLFVWLSPVRANDAMHLSLBEFDNIESQGLIILQ 1430

QY 402 DN-----VLSEI--SQVRLTYLLEDDKFLSPHEMNVHVD 435
 Db 1431 ESFOVWDPKTLIRKGRERHLELFENSIVFSKEVKDSSRSXYLXSKLFTSELGVTETHE 1490
 QY 436 A-----AAFLILSDLRQRPEDMKHYRSVELVQVDEDDAI---YH-- 474
 Db 1491 GDPCKFALWGRTPTSNDKIYLUKASSIENKODMIKHIREVIOERTIHLKGLKEPIHLPK 1550
 QY 475 TSPAL-----GGHTKPODFVILASRRK-----PCDNGDPYVIALRSVTL 513
 Db 1551 TAPATROKGRDGEDLDGSGSGSPOTISIASRSQNTLSDSKSGGCELTVIHDTA 1610
 QY 514 PTHRETPYRGETL-----CSGFLMREGQ 540
 Db 1611 CNSNEL-TIRRGQTEVLERPHDKPDCVLETTR 1644

RESULT 15

US-09-543-681A-5467
 ; Sequence 5467, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETTON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543, 681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 5467
 ; LENGTH: 722
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 US-09-543-681A-5467

Query Match 3.3%; Score 106; DB 4; Length 722;
 Best Local Similarity 18.7%; Pred. No. 0.027;

Matches 118; Conservative 83; Mismatches 171; Indels 260; Gaps 31;

QY 64 LSYGOLLKXIDITACSAERHAGPCVTASMDIYFHTISVGVVNTKAKVNAFNSM 123
 Db 115 VTIG-LMKLYEGAAFIADK--SGAKYIPVRIEGAESYFSRVRKTLINIKAKL----- 163
 QY 124 EVGIGVASEDLCEKQNNVCKALATFVARREITKVKLQITPTEEEKMEHSAABRRM 183
 Db 164 -----FKITIK-VLPAYDLPMPTAKSAERRRL 191
 QY 184 RLIVYADTIDLANCAIQGLDESRCRMVPAEKTRVESYELVPPHANQNTFGQIM 243
 Db 192 S--GEKLDIMDAIMQ----- 206
 QY 244 AMWENAVATTAASRLCAHPTLKALEMFRGSPQVGDRLVKAIVN-----NAPKSM 296
 Db 207 -----ARPVITLYEAF-IQMSQFGRFSPFLISDINMKEDSYQGLKKML 249
 QY 297 EVGVCEAVYQEAETRRRHINSAFMTFVVL-DADDPQLLPWIRPQGD-GERRYREASA 354
 Db 250 GVSRIERTYTERPERIGLLPNTTVYVALFGATMRQRYVAMLTAGSLGVONAKKAAS 309
 QY 355 RKKIRLDKRYIVSCKQTEVPLSVWDPSCNQVYLSTNNVSLKMLVAKDN--WV----- 405
 Db 310 IKTIPTSRQFLKGNLHITPEQTP--DANWIIYL-----EDLKDSITREDKRWIAQHLTP 362
 QY 406 -----LSSEISQVRLTYLED---DKFLS--- 425
 Db 363 HKQMLAQKSTDAVILFTSGSGCTPKGVVSHSHSLANVEQIR--AIADFSPRDKEMAL 420
 QY 426 --FH-MENVVHVA-----AAFLILSDLRQRPEDMKHYRSV-ELVQVDEDDAIHYVTS 476
 Db 421 PLFHAFGLTVSVLAPICLAGRVFLYPSPL-----HYRVVPELY--YDQNTCTVLFGTS 470

QY 477 PALGGHTK---PODFVILASRRKPCDNGDPYVIALRSVTLPTHRETPYRGETLC-SGF 532
 Db 471 TFLGNYGKFAHPDYFARLR-----YVVA-----GAELKSESTR 503
 QY 533 CLMR-----ESDOLTKCQWVSLTELVSASGFYSWGLESRSGRRSDGWNKLAGG 584
 Db 504 VLMQDKFGIRILEGVTGECAPV-VSINVPMASAKPY-----TVGRLLPGMBGRL--- 551
 QY 585 HLSTLKAIPVAKI-----NSRPGYIQ 605
 Db 552 -----IPISGITDGRLOLRGPNIMLGYL 576

Search completed: August 12, 2004, 15:46:22
 Job time : 20 secs

Blank Sheet

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:40:41 ; Search time 13 Seconds

(without alignments)
2431.276 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181

Sequence: 1 MIGNVGNHRLRGASVPSNR.....TLKAIPIVAKINSRFGYLQDT 607

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3181	100.0	607	THEA_HUMAN	Q8WX14 homo sapien
2	2530	79.5	594	THEA_MOUSE	Q8WY99 mus musculu
3	1242.5	39.1	555	CACH_HUMAN	Q8WYX0 homo sapien
4	1240.5	39.0	556	CACH_MOUSE	Q9DBK0 mus musculu
5	1214.5	38.2	556	CACH_RAT	Q99NB7 rattus norv
6	225	7.1	338	BACH_RAT	Q64559 rattus norv
7	223.5	7.0	155	Y654_CHLPM	Q92700 chlamydia p
8	221	6.9	160	BACH_MOUSE	Q91V12 mus musculu
9	214.5	6.7	160	Y535_CHLTR	Q84540 chlamydia t
10	207.5	6.5	159	Y822_CHLMU	Q9P1K7 chlamydia m
11	202	6.4	380	BACH_HUMAN	Q00154 homo sapien
12	177	5.6	174	YKHA_BACST	Q05729 homo sapien
13	175.5	5.5	179	YKHA_HELPY	Q05729 homo sapien
14	145	4.6	439	AC48_MOUSE	Q910X4 mus musculu
15	134.5	4.2	149	YCIA_ZYMO	Q66120 zymomonas m
16	131	4.1	406	CG16_HUMAN	Q9Y305 homo sapien
17	126	4.0	239	ML64_BRARE	Q9DTS4 brachydanio
18	121	3.8	446	ML64_MOUSE	Q61442 mus musculu
19	121	3.8	1711	CHDI_MOUSE	P40201 mus musculu
20	120.5	3.8	233	STR6_MOUSE	P59096 mus musculu
21	118	3.7	445	ML64_HUMAN	Q14649 homo sapien
22	114	3.6	154	YCIA_HAEIN	P44886 haemophilus
23	113.5	3.6	281	STAR_CHICK	Q9D609 gallus gall
24	113	3.6	294	ML64_SALFO	Q14646 homo sapien
25	113	3.6	1709	CHDI_HUMAN	Q89314 buchnera ap
26	107.5	3.4	135	Y254_BUCBP	Q75962 homo sapien
27	106.5	3.3	3038	STR6_HUMAN	P59095 homo sapien
28	105.5	3.3	220	TRIO_HUMAN	P01024 homo sapien
29	104.5	3.3	1663	CO3_HUMAN	P04379 escherichia
30	104	3.3	132	YCIA_ECOLI	P25944 salmonella
31	104	3.3	289	STAR_XENLA	Q8XW22 raietonia s
32	103	3.2	133	YCIA_SALTY	
33	102.5	3.2	985	CAPP_RALSO	

ALIGNMENTS

RESULT 1	ID	THEA_HUMAN	STANDARD:	PRF:	607 AA.	
AC	Q8WX14	075187	Q96D11	Q9H883		P35418 taenia soli
DT	28-FEB-2003	(Rel. 41, Created)				P57362 buchnera ap
DT	28-FEB-2003	(Rel. 41, Last sequence update)				P33731 canis famli
DT	10-OCT-2003	(Rel. 42, Last annotation update)				O76094 homo sapien
DE	Brown fat inducible thioesterase (EC 3.1.2.-)	(BFT)	(Adipose associated thioesterase)			Q98218 molluscum c
GN	THEA OR BFT OR KIAA0707.					Q98218 molluscum c
OS	Homo sapiens (Human)					P28366 bacillus su
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					P42398 buchnera ap
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					Q89541 bradyrhizob
OX	NCBI_TaxID=9606;					Q98ms1 drosophila
RN	[1]					Q99m81 rattus norv
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).					
RX	MEDLINE=21552902; PubMed=11696000;					
RA	Adams S.H., Chui C., Schilbach S.L., Yu X.X., Goddard A.D.,					
RT	Grimaldi J.C., Lee J., Dowd P., Colman S., Lewin D.A.;					
RT	"BFT, a unique acyl-CoA thioesterase induced in thermogenic brown					
RT	adipose tissue: cloning, organization of the human gene and					
RT	assessment of a potential link to obesity.";					
RL	Biochem. J. 360:135-142(2001).					
RL	[2]					
RP	SEQUENCE FROM N.A. (ISOFORM 1).					
RX	TISSUE=Brain;					
RA	MEDLINE=98403880; PubMed=9734811;					
RT	Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,					
RT	Kotani H., Nomura N., Ohara O.;					
RT	"Prediction of the coding sequences of unidentified human genes. X.					
RT	The complete sequences of 100 new cDNA clones from brain which can					
RT	code for large proteins in vitro.";					
RL	DNA Res. 5:169-176(1998).					
RL	[3]					
RP	SEQUENCE FROM N.A. (ISOFORM 2).					
RX	TISSUE=Thyroid;					
RA	Isoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,					
RA	Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,					
RA	Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,					
RA	Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto U.,					
RA	Kakutani A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;					
RT	"NEO human cDNA sequencing project.";					
RT	Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.					
RL	[4]					
RP	SEQUENCE OF 35-255 FROM N.A.					
RX	TISSUE=Skin;					
RA	MEDLINE=22388257; PubMed=12477932;					
RT	Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,					
RT	Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,					
RT	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,					
RT	Stapleton L., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,					
RT	Brownstein M.T., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,					
RT	Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,					
RT	Bosak S.A., McEwan P.J., McKernan K.D., Malek J.A., Gunaratne P.H.,					

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Whiting J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schenker A., Schein J.B., Jones S.J.M., Maira M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Has acyl-CoA thioesterase activity towards medium (C12)
 CC and long-chain (C18) fatty acyl-CoA substrates.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=BFIT1;
 CC IsoId=08MX14-1; Sequence=Displayed;
 CC Name=2; Synonyms=BFIT2;
 CC IsoId=08MX14-2; Sequence=VSP 000160;
 CC -1- TISSUE SPECIFICITY: Isoform 1 is predominantly expressed in
 CC skeletal muscle, liver, testis, stomach, spleen, lung and brain.
 CC Isoform 2 is predominantly expressed in kidney, uterus, hibernoma
 CC and white adipose tissue.
 CC -1- INDUCTION: By cold exposure and repressed by heat exposure.
 CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
 CC -1- SIMILARITY: Contains 1 START domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF416921; AAL40937.1; -;
 DR EMBL; AF416922; AAL40938.1; -;
 DR EMBL; AB014607; BAA31682.1; AUT_INIT.
 DR EMBL; AK023937; BAA14734.1; -;
 DR EMBL; BC001517; AAH01517.1; -;
 DR GeneW; HGNC:18156; THEA.
 DR MIM; 606803; -;
 DR GO; GO:0005737; C:cytoplasm; IC.
 DR GO; GO:0016291; F:acyl-CoA thioesterase activity; TAS.
 DR GO; GO:0006631; P:fatty acid metabolism; NAS.
 DR GO; GO:0007242; P:intracellular signaling cascade; NAS.
 DR GO; GO:0009266; P:response to temperature; ISS.
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_supe.
 DR Pfam; PF03061; 4HBT; 2.
 DR Pfam; PF01852; START; 1.
 DR SMART; SMO0234; START; 1.
 DR PROSITE; PS00848; START; 1.
 KM Hydrolyase; Serine esterase; Repeat; Alternative splicing.
 PT DOMAIN 29 164 ACYL-CoA HYDROLASE 1.
 PT DOMAIN 205 336 ACYL-CoA HYDROLASE 2.
 PT DOMAIN 375 585 START.
 VARSPLIC 544 607
 CCWVSILTEIVSASGFYSWGIESRSKGRSDGNMKGKLAG
 HSTLTKAIPVAKINSRFGYLODT -> VSYNQATPGVINY
 VTTNVAIGLSSEFYTFKACEGFLDNRNDLAPLSLOT (in
 isoform 2).
 /FTID-VSP 000160.
 S -> R (IN REF. 4).
 FT CONFLICT 255 255
 FT SEQUENCE 607 AA; 68492 MW; 12F2BCBA8AAC18BC CRC64;
 Query Match 100.0%; Score 3181; DB 1; Length 607;
 Best Local Similarity 100.0%; Pred. No. 6,2e-244;
 Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 RGEISVQGLKMTDTTACTLSAERHAGGCVTASMDITYFEHTISVGVNVIKAKVNAFN 120
 DB 61 RGEISVQGLKMTDTTACTLSAERHAGGCVTASMDITYFEHTISVGVNVIKAKVNAFN 120
 QY 121 SSMEVGIOVASSEDICSEKQNVNVCALATPVAREITVKLKQITPRTTEERMEHVAER 180
 DB 121 SSMEVGIOVASSEDICSEKQNVNVCALATPVAREITVKLKQITPRTTEERMEHVAER 180
 QY 181 RRMELVYADDTIKOLLANCAIOGDLSESDCSRMVAEKTRVESVTLVPPHANHGNTFGG 240
 DB 181 RRMELVYADDTIKOLLANCAIOGDLSESDCSRMVAEKTRVESVTLVPPHANHGNTFGG 240
 QY 241 QIYAMMENAVATIASRCLRAHPTLKALEMFEFRSPQVGDRLVYKALVNNAFKSMVEGV 300
 DB 241 QIYAMMENAVATIASRCLRAHPTLKALEMFEFRSPQVGDRLVYKALVNNAFKSMVEGV 300
 QY 301 CVEAYRQEAETHRRHINSAPMTFVLDADDPQOLLPIWIRPQDGERRYRASARKIRL 360
 DB 301 CVEAYRQEAETHRRHINSAPMTFVLDADDPQOLLPIWIRPQDGERRYRASARKIRL 360
 QY 361 DRKTVSCQKQTEVPLSPWDPNSQVYLSYNNVSLKMLVAKDNWVLSSEISOVRLTYLED 420
 DB 361 DRKTVSCQKQTEVPLSPWDPNSQVYLSYNNVSLKMLVAKDNWVLSSEISOVRLTYLED 420
 QY 421 DKFLSPHMEVHVADAQAFLILSDLRQRPMDKHYRSVELVOQVEDDAIYHVTSPALG 480
 DB 421 DKFLSPHMEVHVADAQAFLILSDLRQRPMDKHYRSVELVOQVEDDAIYHVTSPALG 480
 QY 481 GHTXPQDFVLIASRRKPCDNDPYVIALRSVTLPTHRETPEYRGETLCSGFCILMBEDQ 540
 DB 481 GHTXPQDFVLIASRRKPCDNDPYVIALRSVTLPTHRETPEYRGETLCSGFCILMBEDQ 540
 QY 541 LTRKQWRVSLTELVSASGFYSWGIESRSKGRSDGNMKGKLAGHLSLTAKAIPVAKINSR 600
 DB 541 LTRKQWRVSLTELVSASGFYSWGIESRSKGRSDGNMKGKLAGHLSLTAKAIPVAKINSR 600
 QY 601 FGYLODT 607
 DB 601 FGYLODT 607
 RESULT 2
 THEA_MOUSE
 ID THEA_MOUSE STANDARD; PRT; 594 AA.
 AC Q8VHQ9; 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Brown fat inducible thioesterase (EC 3.1.2.-) (BFIT) (Adipose
 DE associated thioesterase).
 GN THEA OR BFIT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=21552902; PubMed=11696000;
 RA Adams S.H., Chui C., Schilbach S.U., Yu X.X., Goddard A.D.,
 RA Grimaldi J.C., Lee J., Dowd P., Colman S., Lewin D.A.,
 RT "BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown
 RT adipose tissue: cloning, organization of the human gene and
 RT assessment of a potential link to obesity.";
 RT Biochem. J. 360:135-142(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; Tissue=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Struhsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen G.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marsden K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Caminici P., Prange C.,
 RA Rata S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Wotley K.C., Hale S., Garcia A.M., Gay L.U., Holyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.U.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Has acyl-CoA thioesterase activity towards medium (C12)
 CC and long-chain (C18) fatty acyl-CoA substrates.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- INDUCTION: By cold exposure and repressed by heat exposure.
 CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
 CC -1- SIMILARITY: Contains 1 START domain.

Db 240 GGOIMAMMENVAITIAASRLCHAPHTLKAIMEFHFRGSPQVGDRLVTLKAIYNNAPKSHMEV 299
 Qy 299 GVCVEAARQAEFRHRIINSAFTFVYLDADDQDOLLPWIRPOGDERRRREASARKKI 358
 Db 300 GVCVEARQAEFRHRIINSAFTFVYLDADDQDOLLPWIRPOGDERRRREASARKKI 359
 Qy 359 RLDRKTYVSCQTEVPLVPWDSNQYLYSYNNVSLKMLVAKDNWLVSEISQVRYTL 418
 Db 360 RLDRKTYVSCQTEVPLVPWDSNQYLYSYNNVSLKMLVAKDNWLVSEISQVRYTL 419
 Qy 419 EDDKFLSFHEHRYVYVDAQAFLSLDRQPEWDRKYSVELVQYVDEDDAIYHTSPA 478
 Db 420 EED-FLSFHEHRYVYVDAQAFLSLDRQPEWDRKYSVELVQYVDEDDAIYHTSPA 478
 Qy 479 LGGHTXPQDQVVIILASRKKPCDNGPYVIALRSVTLPHRTPEYRGETTCSGFCIMREG 538
 Db 479 LSGHTXPQDQVVIILASRKKPCDNGPYVIALRSVTLPHRTPEYRGETTCSGFCIMREG 538
 Qy 539 DQITKCCVAVSVLTETVSAAGFYSGWLSKSGKRRSDGNGKLAGHILTLKAIPAKIN 598
 Db 539 DQITKCCVAVSVLTETVSAAGFYSGWLSKSGKRRSDGNGKLAGHILTLKAIPAKIN 598
 Qy 599 SR 600
 Db 583 NR 584

RESULT 3
 CACH HUMAN STANDARD; PRT; 555 AA.
 ID CACH HUMAN
 AC Q8WYK0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytoplasmic acetyl-CoA hydrolase 1 (BC 3.1.2.1) (CACH-1) (NCACH-1).
 GN CACH1 OR CACH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Suematsu N., Okamoto K., Isohashi F.,
 RT "Cytosolic acetyl-CoA hydrolase.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Hydrolyzes acetyl-CoA to acetate and CoA (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O = CoA + acetate.
 CC -1- PATHWAY: Probably involved in fatty acid metabolism.
 CC -1- SUBUNIT: Homodimer or homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
 CC -1- SIMILARITY: Contains 1 START domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF416923; AA40939.1; -;
 DR EMBL; BC042492; AA42492.1; -;
 DR MGI; 1913736; Thea.
 DR GO; GO:0005737; C:cytoplasm; IC.
 DR GO; GO:0016291; F:acyl-CoA thioesterase activity; TAS.
 DR GO; GO:0006631; P:fatty acid metabolism; NAS.
 DR GO; GO:0007242; P:intracellular signaling cascade; NAS.
 DR GO; GO:0009266; P:response to temperature; IEI.
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR Pfam; PF01852; START; 1.
 DR SMART; SM00234; START; 1.
 DR PROSITE; PS00848; START; 1.
 FM Hydrolyase; Serine esterase; Repeat.
 FT DOMAIN 31 166 ACYL-CoA HYDROLASE 1.
 FT DOMAIN 201 337 ACYL-CoA HYDROLASE 2.
 FT DOMAIN 370 582 START.
 SQ SEQUENCE 594 AA; 67354 MW; 9A51D3BCF38870C9 CRC64;

Query Match 79.5%; Score 2530; DB 1; Length 594;
 Best Local Similarity 81.9%; Pred. No. 2e-192;
 Matches 493; Conservative 35; Mismatches 54; Indels 20; Gaps 6;

Qy 1 MIGNVGNHLRGLAVSFNSRTSRKSAI--RAGNDSAMADEGGRNPTEVMSQVLVPCHT 58
 Db 1 MIGNVGNHLRGLAVSFNSRTSRKSAI--RAGNDSAMADEGGRNPTEVMSQVLVPCHT 60
 Qy 59 NQGEHLSVGLLKWIDTTACLSARRHAGCCPVTSMDIYFEHTISVGQVNVTKAKYNNR 118
 Db 61 NHRGEHLSIGGLLKWIDTTACLSARRHAGCCPVTSMDIYFEHTISVGQVNVTKAKYNNR 120
 Qy 119 FNSMSEVGIQVASEDCSEKQNVVCKALAFVVARRETTKKLCQITRTREEEKHESVVA 178
 Db 121 FNSMSEVGIQVASEDCSEKQNVVCKALAFVVARRETTKKLCQITRTREEEKHESVVA 180
 Qy 179 ERRRRRLVYADITKDLILANCAIOGDLSEKSDSRNVPAKTRVSEVELVPPHANHOGNTF 238
 Db 181 ERRRRRLVYADITKDLILTHCVIQDDLD-KDCSNMVPAEKTRVSEVELVPPHANHOGNTF 239
 Qy 239 GGOIMAMMENVAITIAASRLCHAPHTLKAIMEFHFRGSPQVGDRLVTLKAIYNNAPKSHMEV 298

DR EMBL; AB078619; BAB84022.1; -;
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PF03061; 4HBT; 2.
 DR Pfam; PF01852; START; 1.
 DR PROSITE; PS00848; START; 1.
 FM Hydrolyase; Serine esterase; Fatty acid metabolism; Repeat.
 FT DOMAIN 1 127 ACYL-CoA HYDROLASE 1.
 FT DOMAIN 165 301 ACYL-CoA HYDROLASE 2.

```

FT DOMAIN 340 549 START.
SQ SEQUENCE 555 AA; 62034 MM; 707560D55504732C CEC64;
Query Match 39.1%; Score 1242.5; DB 1; Length 555;
Best Local Similarity 52.2%; Pred. No. 1.5e-90;
Matches 260; Conservative 82; Mismatches 143; Indels 13; Gaps 7;

QY 44 PTEVOMGLVPCHTNORGEISVGLKWDITTAQCLASERRHAGPCVTAAMDLYEHTI 103
DB 6 PGEVWMSOALQPAATYATGAGELISAGOLKMDITTAQCLAEKRAAGVSAVSDDIQFETA 65
QY 104 SVGVVNNIKAKVNAFANSSMEVGIQVASEDLCEKQNMVACALATFVAR--REITKYKL 160
DB 66 RVGQVITIKAVTAFASTSMELISKVMVQDMLTGTEKVSVAFTFVAKPKGKE--KHHL 123
QY 161 KQITPRTEEEKHESVAERRRRRLVADTKDLANCAIQGDLESDSCSMVPAEKTRV 220
DB 124 KPVTLTLEQDHVEHNLAEERRKRLQHEDTFNNTLMKESKFDLILFDEEGAVSTRTGSV 183
QY 221 ESVELVLPPHANNHGNFTGQIMAMMENVAITLASRLCRAPPTLKALMEHFRGPGVQVD 280
DB 184 QSIELVLPPHANNHGNFTGQIMAMMENVAITLASRLCAHPFLKSDMEKFRGPGSTVG 243
QY 281 RLVLKATVNNAFKHSMEVGVCEAV--RQEAETHRRHINSAPMTFVVDADDDQQLT--P 336
DB 244 RLVTALVNNFTFCVCEVGVAVPAFCQEMAEGRHINSAPFLVYNA--ADDKENTLITFP 301
QY 337 WIRPOGDEGERRYPASARKKIRLDKRTVSCQKTEVPLVSPWDSNOVLYSYNNVSLK 396
DB 302 RIQISKDKDFRRYGAIAKRRIRLRGRKYVISHKE--EYPLCHIMWIDSQASLSDSVEALK 360
QY 397 MLVAKDMNVLSSEISQVRLTLEDDKELSPFEMEMVYHNDQAQFLISDLKORPEWDEKY 456
DB 361 KLAARQGEVSTVEKRIYTLREHDVLSVWEKRVGSPAHARLKLSDFTKRLPMDDHF 420
QY 457 RSVELVQVDDDAIYVHTSPALGGHTKQDFVILASRRKCDNDGPVIALRSVLTPTH 516
DB 421 VSCVIVDMSDDQQLYHTCPIL--NDKPKDLVYVSRKRLKQNTYVAVKSVTLPSV 479
QY 517 RETPEYRERFGLCSGFLC 534
DB 480 PPSQYIRSEITICAGFLI 497

RESULT 4
CACH MOUSE STANDARD; PRT; 556 AA.
AC Q9DBK0; Q8R108;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cytoplasmic acetyl-CoA hydrolase 1 (EC 3.1.2.1) (CACH-1) (mCACH-1).
GN CACH1 OR CACH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22430267; PubMed=12545200;
RA Sumatru N., Okamoto K., Iwashita F.;
RA "Mouse cytosolic acetyl-CoA hydrolase, a novel candidate for a key
RT enzyme involved in fat metabolism: cDNA cloning, sequencing and
RT functional expression."
RL Acta Biochim. Pol. 49:937-945 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aikawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,

```

```

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batola S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochenush J.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Nomberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
RA Wyszewski A., Yoshida K., Hasegawa Y., Kawaji H., Kohlschki S.,
RA Watanabe Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE OF 14-556 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepheon M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bata S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halik S.W.,
RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boufard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RC -1- FUNCTION: Hydrolyzes acetyl-CoA to acetate and CoA (by
CC similarity).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O = CoA + acetate.
CC -1- PATHWAY: Probably involved in fatty acid metabolism.
CC -1- SUBUNIT: Homodimer or homotetramer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
CC -1- SIMILARITY: Contains 1 START domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AB078618; BAB84021.1; -
DR EMBL, AK004905; BAB23658.1; -
DR EMBL, BC025852; AAB25852.1; -
DR GDB, MGI:1921406; Cach.
DR GO, GO:0003986; F:acetyl-CoA hydrolase activity; IDA.
DR GO, GO:0005524; F:ATP binding; IDA.
DR GO, GO:0006084; F:acetyl-CoA metabolism; IDA.
DR InterPro, IPR002913; START.
DR InterPro, IPR006683; Thioester_suff.
DR Pfam, PF03061; 4HBT; 2.
DR Pfam, PF01852; START; 1.
DR PROSITE, PS00848; START; 1.
KW Hydrolase, Serine esterase, Fatty acid metabolism, Repeat.
FT DOMAIN 1 128 ACYL-CoA HYDROLASE 1.
FT DOMAIN 162 302 ACYL-CoA HYDROLASE 2.

```


FT DOMAIN 327 536 START
 S0 SEQUENCE 556 AA; 61761 MW; 9FE9C487B0CB812 CRC64;
 Query Match 39.0%; Score 1240.5; DB 1; Length 556;
 Best Local Similarity 50.8%; Pred. No. 2,1e-99;
 Matches 253; Conservative 93; Mismatches 139; Indels 13; Gaps 7;

QY 44 PTEVQNSQVLPCHTNORGLSVGQILKMTDTTACLSAERHAGCPCTASMDIYFHTI 103
 Bb 7 PGEVMSQALQPAHADSRGELISAGQILKMTDTTACLSAERHAGISCTASMDILFEDTA 66
 QY 104 SVGVNVKAKVRAFPSSMEVGIQVASEDLCSKMNCKALATFPAR--REITVVKL 160
 Db 67 RIQGITITRAKVTAFSTSMELISKVIVQDKFTGIQILCLVASTFAKVGKE--KVHL 124
 QY 161 KQITPTEERKMEHSAVERRRMRLVYADITKDLANCAIQGLDSRDMYPAEKTRV 220
 Db 125 KPVLTQTEQGVHNLASERKXVRLQHEHTFNNIMKSSRPSDSICNEEGTATMTGTSV 184
 QY 221 ESVELVLPPIRANHQNTFGGQIMAMMENVAITIASRLCRAHPTLKALIEHFRGPGVGD 280
 Db 185 QSIETLVLPPIRANHGNFTFGGQIMAMMETVATISASRLCHGHPFLKSDVMEKFRGPSTVGD 244
 QY 281 RLVLKALIVNNAFQSHMEVGVCEAV--ROEAEHTRHINSAPMTFVVLDDDDPOLL--P 336
 Db 245 RLVPFSAIVNNTFQNSVEGVVRAFDCEWAEQGGRIHNSAFILYNV--DDDEKILTFP 302
 QY 337 WIRPODGERRRYREASARKKIRLDKRYISCKQTEVELSVMPDPSNOVYLSTNNVSLK 396
 Db 303 RIQPIKSDPFRRYOGAIIARRIRLGRKYVISHKK--EVLPLQWMDISKSGISLNTVVALK 361
 QY 397 MLYAKDNWVLSSISQVRLTYLLEDDKFLSFHMEVNVHDAQAFLISDLRQREPMKHY 456
 Db 362 NLASKSGWEITTLTEKIKYITLLEQDAISVKEKLVGSPAHIAVHLLSDLTKEPLMDPHY 421
 QY 457 RSEVLVQVVEDDAIYHVTSPALGHTKPODFVILASRRKPCNGDPYVALASVTLPTH 516
 Db 422 ISCEVLDQVEDDQIYITTSVNG--DKPDQFVVLVSRKPLKDNNTYVALRSVVLPSV 480
 QY 517 RETPEYRGETLCSGFL 534
 Db 481 PSSQYIRSGVTCAGFLI 498

RESULT 5
 CACH RAT STANDARD; PRT; 556 AA.
 ID CACH RAT
 AC 099NE7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DE 10-OCT-2003 (Rel. 42, last annotation update)
 GN Cytoplasmic acetyl-CoA hydrolase 1 (EC 3.1.2.1) (CACH-1) (ZACH).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 161-174 AND 352-364.
 RC STRAIN=Donryu; TISSUE=liver;
 RX MEDLINE=2123026; PubMed=11322891;
 RA Suematsu N., Okamoto K., Shibata K., Nakanishi Y., Isohashi F.;
 RT Molecular cloning and functional expression of rat liver cytosolic
 RT acetyl-CoA hydrolase.";
 RL Eur. J. Biochem. 268:2700-2709(2001).
 CC -1- FUNCTION: Hydrolyzes acetyl-CoA to acetate and CoA.
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O = CoA + acetate.
 CC -1- ENZYME REGULATION: Allosterically regulated by ATP (activator) and
 CC ADP (inhibitor).
 CC -1- PATHWAY: Probably involved in fatty acid metabolism.
 CC -1- SUBUNIT: Active homodimer or homotrimer at room temperature and
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- INDUCTION: By 2-(p-chlorophenoxy)isobutyric acid (CPIB).
 CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
 CC -1- SIMILARITY: Contains 1 START domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AB040609; BAB39852.1; -;
 DR InterPro; IPR002913; START.
 DR InterPro; IPR006683; Thioestr_supf.
 DR Pfam; PFO3061; 4HT; 2.
 DR Pfam; PFO1852; START; 1.
 DR PROSITE; PS50848; START; 1.
 KW Hydrolase; Serine esterase; Fatty acid metabolism; Repeat;
 KW Allosteric enzyme.
 FT DOMAIN 1 128 ACYL-COA HYDROLASE 1.
 FT DOMAIN 163 302 ACYL-COA HYDROLASE 2.
 FT DOMAIN 341 550 START
 S0 SEQUENCE 556 AA; 62018 MW; 81E592F066AB0C9E CRC64;
 Query Match 38.2%; Score 1214.5; DB 1; Length 556;
 Best Local Similarity 49.8%; Pred. No. 2,4e-88;
 Matches 248; Conservative 92; Mismatches 145; Indels 13; Gaps 7;

QY 44 PTEVQNSQVLPCHTNORGLSVGQILKMTDTTACLSAERHAGCPCTASMDIYFHTI 103
 Bb 7 PGEVMSQALQPAHADSRGELISAGQILKMTDTTACLSAERHAGISCTASMDILFEDTA 66
 QY 104 SVGVNVKAKVRAFPSSMEVGIQVASEDLCSKMNCKALATFPAR--REITVVKL 160
 Db 67 RIQGITITRAKVTAFSTSMELISKVIVQDKFTGIQILCLVASTFAKVGKE--KVHL 124
 QY 161 KQITPTEERKMEHSAVERRRMRLVYADITKDLANCAIQGLDSRDMYPAEKTRV 220
 Db 125 KPVLTQTEQGVHNLASERKXVRLQHEHTFNNIMKSSNMLRDPVNCNEEGTATMTGTSV 184
 QY 221 ESVELVLPPIRANHQNTFGGQIMAMMENVAITIASRLCRAHPTLKALIEHFRGPGVGD 280
 Db 185 QSIETLVLPPIRANHGNFTFGGQIMAMMETVATISASRLCHGHPFLKSDVMEKFRGPSTVGD 244
 QY 281 RLVLKALIVNNAFQSHMEVGVCEAV--ROEAEHTRHINSAPMTFVVLDDDDPOLL--P 336
 Db 245 RLVPFSAIVNNTFQNSVEGVVRAFDCEWAEQGGRIHNSAFILYNV--DDDEKILTFP 302
 QY 337 WIRPODGERRRYREASARKKIRLDKRYISCKQTEVELSVMPDPSNOVYLSTNNVSLK 396
 Db 303 RIQPIKSDPFRRYOGAIIARRIRLGRKYVISHKK--EVLPLQWMDISKSGISLNTVVALK 361
 QY 397 MLYAKDNWVLSSISQVRLTYLLEDDKFLSFHMEVNVHDAQAFLISDLRQREPMKHY 456
 Db 362 NLASKSGWEITTLTEKIKYITLLEQDAISVKEKLVGSPAHIAVHLLSDLTKEPLMDPHY 421
 QY 457 RSEVLVQVVEDDAIYHVTSPALGHTKPODFVILASRRKPCNGDPYVALASVTLPTH 516
 Db 422 ISCEVLDQVEDDQIYITTSVNG--DKPDQFVVLVSRKPLKDNNTYVALRSVVLPSV 480
 QY 517 RETPEYRGETLCSGFL 534
 Db 481 PSSQYIRSGVTCAGFLI 498

RESULT 6
 BACH RAT STANDARD; PRT; 338 AA.
 ID BACH RAT
 AC Q64559; O08652; O09041;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 15-MAR-2004 (Rel. 43, last annotation update)


```

RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae A39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Onuchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA."
RL Nucleic Acids Res. 28:2311-2314 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TW-183;
RA Gang M.M., Schumacher A., Mehlendorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB001648; AAD18793.1; -
DR EMBL: AB002172; AAF37977.1; -
DR EMBL: AB002547; BAA98861.1; -
DR EMBL: AB017159; AAP98609.1; -
DR PIR: C86572; C86572.
DR PIR: G72051; G72051.
DR TIGR: CP0093; -
DR InterPro: IPR006683; Thioestr_supf.
DR Pfam: PF03061; 4HBT; 1.
KW Hypothetical protein; Hydrolase; Complete proteome.
SQ SEQUENCE 155 AA; 17883 MW; FE1483EA30323AA CRC64;

Query Match          7.0%; Score 223.5; DB 1; Length 155;
Best Local Similarity 35.5%; Pred. No. 1.1e-10;
Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;

QY 226 VLPPHANGQNTFGGOIMAWMENAVTIAASRLGRAHPTLCAIMFHRGSPGSDRLVYK 285
DB 19 IFENDINANTVFGGLMSLDRLALVVAERHTESCVTVAFVDALRYADAYMGENTLCK 78
QY 286 AIVNNAFKHSMVEVCVEA---YRQEAETRRHINSAMFTFVVLADDDQPOLPWIRPCP 342
DB 79 AAVNRRTWRTSLFVGKVAWENIKQE---RHHTSAVFVFVAWENQGIPIVHQIVPET 134
QY 343 GDGERYRERSAKKRIIDRK 363
DB 135 PEKKRYNEADRRRRQRLLEIK 155

RESULT 8
BACH_MOUSE
ID BACH_MOUSE STANDARD; PRT; 381 AA.
AC Q91V12;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytosolic acyl coenzyme A thioester hydrolase (EC 3.1.2.2) (Long chain
DE acyl-CoA thioester hydrolase) (CTE-II) (Brain acyl-CoA hydrolase).
GN BACH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC STRAIN=ICR; TISSUE=Brain;
RA Takagi-Sakuma M., Kuramochi Y., Kitchara M., Emori R., Kuroda J.,
RA Hiratsuka K., Nagae Y., Watanabe T., Suga T., Unji Y.;
RT "Mouse brain long-chain acyl-CoA hydrolase."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS B AND C).
RC STRAIN=ICR; TISSUE=Brain;
RA Yamada J., Kuramochi Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heile F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshimiki S., Carninci P., Prange C.,
RA Kana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting M., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson W.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smillie D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -1- FUNCTION: May play an important physiological function in brain.
CC May play a regulatory role by modulating the cellular levels of
CC fatty acyl-CoA ligands for certain transcription factors as well
CC as the substrates for fatty acid metabolizing enzymes,
CC contributing to lipid homeostasis. Has broad specificity, active
CC towards fatty acyl-CoAs with chain-lengths of C8-C18. Has a
CC maximal activity toward palmitoyl-CoA.
CC -1- CATALYTIC ACTIVITY: Palmitoyl-CoA + H(2)O = CoA + palmitate.
CC -1- SUBUNIT: Homodimer (Probable).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=B; Synonyms=mbACHb;
CC IsoId=Q91V12-1; Sequence=Displayed;
CC Name=A; Synonyms=mbACHa;
CC IsoId=Q91V12-2; Sequence=VSP_000158;
CC Name=C; Synonyms=mbACHc;
CC IsoId=Q91V12-3; Sequence=VSP_000157;
CC -1- SIMILARITY: Contains 2 acyl coenzyme A hydrolase domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB049821; BAB61731.1; -
DR EMBL: AB088411; BAC20217.1; -
DR EMBL: AB088412; BAC20218.1; -
DR EMBL: BC013507; AAI13507.1; -
DR MGD: MGI:1917275; Bach.
DR GO: GO:0005829; C:cytosol; IDA.
DR GO: GO:0009062; P:fatty acid catabolism; IDA.
DR InterPro: IPR008994; Nucleic_acid_OB.
DR InterPro: IPR006683; Thioestr_supf.

```

DR Pfam; PF03061; 4HBT; 2
 KW Hydrolyase; Serine esterase; Repeat; Alternative splicing.
 FT DOMAIN 44 178 ACYL-COA HYDROLASE 1.
 FT DOMAIN 210 346 ACYL-COA HYDROLASE 2.
 FT VARSPPLIC 1 59 MKLVGTLRIMEVGRQVAFSLTPGQSGIKRTFMAMRA
 VRTADHOKLGHCVTMG -> MTLTRHALRVLRKREVE
 AYLRKVKO (in isoform C).
 FT FTID=VSP_000157.
 FT VARSPPLIC 1 58 MKLVGTLRIMEVGRQVAFSLTPGQSGIKRTFMAMRA
 VRTADHOKLGHCVTMG -> MSGPTDTPAIDIC (in
 isoform A).
 FT FTID=VSP_000158.
 FT SEQUENCE 381 AA; 42537 MW; 813852DBB6834C4 CRC64;
 Query Match 6.9%; Score 221; DB 1; Length 381;
 Best Local Similarity 22.6%; Pred. No. 6e-10;
 Matches 74; Conservative 64; Mismatches 154; Indels 36; Gaps 7;
 QY 47 VQMSQVLPCHTNGRGELSVGLKMTDTACLSAERAG-----CPCTASMDIYFE 100
 DB 55 VTMGRIMPDADANVAGVHGSTILKMLEAGAITSTRHNSQNGERCVAALARYBRIDFL 114
 QY 101 HTLSVGQVNIKAKVNAFNSMVEGIVASEDLCEKQNVCKALATFV--ARRBITKV 158
 DB 115 SPMGIGFAVHASAETITTSKSHVEQVHVMSENILTGKTLTKATLMYVPLSKNDVKV 174
 QY 159 ----KLQITRTBEKKMEHSVAAERRRMELV--ADITKOLLNCAIQGLSEGRDSRM 212
 DB 175 LEVPEIYVLRQOEHEGKRKYAOKLEMETKMRNGDIDVQPLN----- 218
 QY 213 VPARKT-----RVESAEVLPRPHANQGTFGQIMAMENAVATIAASRLCAHPTLKAE 268
 DB 219 -PEPVNTSYSSSLIHVPSDCTLHGVRHGVTKMLMDRYAGVIAAHCKNTIVTASVD 277
 QY 269 MFHFRGSPQVDRVLVLAIVNAFHSMEVGVCEAYEQEAEHRRH-INSAPMTFVLVD 327
 DB 278 AINPHDKIRKGCVALITISGRMTFTSKNSWEIEVLVDADVNSQGRYPAASAFETYELN 337
 QY 328 ADDQQLPMTIRPQDGERRYREASAR 355
 DB 338 QEGKMPVPQVLVPEDEKRRFEKGGR 365
 RESULT 9
 Y535 CHLUTR STANDARD; PRT; 160 AA.
 ID Y535 CHLUTR
 AC 084540;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative acyl-CoA thioester hydrolase CT535 (EC 3.1.2.-).
 OS Chlamydia trachomatis.
 CC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=813;
 CC NCR1_TaxID=813;
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kallman S., Lamell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis";
 RL Science 282:754-759(1998).
 CC -I- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC EMBL; AE001324; AAC68137.1; -.
 DR PIR; F71502; F71502. Thioestr_suff.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 160 AA; 18526 MW; 90B81F56F13BA6F2 CRC64;
 Query Match 6.7%; Score 214.5; DB 1; Length 160;
 Best Local Similarity 34.1%; Pred. No. 5.8e-10;
 Matches 47; Conservative 30; Mismatches 54; Indels 7; Gaps 2;
 QY 226 VLPRHANQGTFGQIMAMENAVATIAASRLCAHPTLKAEIMFHFRGSPQVGRVLK 285
 DB 25 IFPNDLMTNNTIFGGLMSLIDRLALVVAEHGCSICTATVADAMRFPAPVMEGLICC 84
 QY 286 AIVNAFHSMEVGVCEA--YRQEAETHRRHNSAFMTFVLADQQLPMTIRPQ 342
 DB 85 ASVNRSMRTSLSEGVKVAENITKOE--HRHITSAFTEFVAVDKNNSPVEVPELPES 140
 QY 343 GDGERRRREASARKIRL 360
 DB 141 QEIRRRREADQRRALRL 158
 RESULT 10
 Y822 CHLMTU STANDARD; PRT; 159 AA.
 ID Y822 CHLMTU
 AC Q9PJX7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative acyl-CoA thioester hydrolase TC0822 (EC 3.1.2.-).
 OS Chlamydia muridarum.
 CC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=83560;
 CC NCR1_TaxID=83560;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mopn / Nigg;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uteerback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClary G., Salzberg S.L.,
 RA Eisen J., Fraser C.W.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -I- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC EMBL; AE002347; AAF39624.1; -.
 DR PIR; C81662; C81662.
 DR TIGR; TC0822; -.
 DR InterPro; IPR006683; Thioestr_suff.
 DR Pfam; PF03061; 4HBT; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 159 AA; 18248 MW; EFNAD7BDCB692B11 CRC64;
 Query Match 6.5%; Score 207.5; DB 1; Length 159;
 Best Local Similarity 33.3%; Pred. No. 2.1e-09;
 Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;
 QY 226 VLPRHANQGTFGQIMAMENAVATIAASRLCAHPTLKAEIMFHFRGSPQVGRVLK 285

Db 24 IFPNDLNTNNNTTIFGGLLSMLDLDELAIWAERFCEISICVTALDAVFAFAPAYMOENLICC 83

Oy 286 AIYVNAKHSMEVGVCEA---YRGAERHRHINSAPMTFVLVDADQPOLMIRPQP 342

Db 84 ASVNRSMRTSLFEGVKKWAMENIKOE---QRHITSAPYTFVAVDENNSPVEIPELVET 135

Oy 343 GDGGRRYEASARKIRL 360

Db 140 QEIRREFHEADQRRASRL 157

RESULT 11

BACH HUMAN

ID BACH HUMAN STANDARD: PRT: 380 AA.

AC 000154: 043703: QJUM06: Q9Y539: Q9Y540;

DT 01-NOV-1997 (Ref. 35, Created)

DT 18-FEB-2003 (Ref. 41, Last sequence update)

DT 10-OCT-2003 (Ref. 42, Last annotation update)

DE CytoSolic acyl Coenzyme A thioester hydrolase (EC 3.1.2.2) (Long chain acyl-CoA thioester hydrolase) (CTE-II) (Brain acyl-CoA hydrolase).

GN BACH.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

CC NCBI_TaxID=9606;

RP [1]

RP SEQUENCE FROM N.A. (ISOFORM 4), AND CHARACTERIZATION.

RC TISSUE=Brain;

FX MEDLINE=20047053; PubMed=10578051;

RA Yamada J., Kurata A., Hirata M., Ianiuchi T., Takama H., Furinata T., Shiratori K., Iida N., Takegi-Sakuma M., Watanabe T., Kurosaki K., Endo T., Suga T.;

RT "Purification, molecular cloning, and genomic organization of human brain long-chain acyl-CoA hydrolase.";

RL J. Biochem. 126:1013-1019(1999).

RP [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Hippocampus;

RA Hajira A.K., Uhler M.D., Iarkins L.K.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RP [3]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 5 AND 6).

RA Yamada J., Kuramochi Y.;

RT "Human brain acyl-CoA hydrolase isoforms.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

RP [4]

RP SEQUENCE FROM N.A. (ISOFORM 4).

RC TISSUE=Brain;

FX MEDLINE=22388257; PubMed=12477932;

RA Klausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Stausberg R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Bueltow K.H., Scheffer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Rana S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gamaratte P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hallyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Keltman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Rattiner J.E., Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP [5]

RP SEQUENCE OF 1-247 FROM N.A.

RA Howden P.,;

[illegible]

```

FT VARSPLIC 287 288 GC -> AP (in isoform 2).
FT VARSPLIC 289 380 Missing (in isoform 2).
FT CONFLICT 371 380 /FTID=VSP_000136.
FT SEQUENCE 380 AA; 41796 MW; BDD75D62A60095BC CRC64;
SQ
Query Match
Best Local Similarity 21.9%; Pred. No. 1.9e-08;
Matches 71; Conservative 61; Mismatches 156; Indels 36; Gaps 6;
QY 51 QLVLPCHTNGRGLSVGQLTKMIDTTCASAEHRHAG-----CPCTASMDIYFHTIS 104
Db 58 RIMPPDDANVANGNHCITLKMIEAGAIISTRHNSNGRCVAAALAVERTDPLSMC 117
QY 105 VGQVNNIKAKYNNRANFNSMEVGIQVASEDLCEKQMNCKALAFVAREITTKVLTQIT 164
Db 118 IGEVAHVAEITLYSKSHVEVQVNMSENILTGAKKLTNKAITLWVPLSLKMKVDKLEVP 177
QY 165 P-----RTEREKXHSVAERBRRLVY--ADTIKDLNCAIQGLSDSCRMVPAE 216
Db 178 PVTYSRQGEERGRYRQAQLERETKWRNGDIVQVLA-----PEP 220
QY 217 KT-----RVESVELVLPFHNHQNTFGQIMAMMENVAITIASRLCRAPTLKAIEMFHF 272
Db 221 NTVSYSQSLHLIVGSPDCTLHGFGVGMKMDVAGIYAARCKNTNITASVDALNF 280
QY 273 RGSQVQDRLVYLKATVNNAPFHSMEVGVCEAYNROEAETRRH--INSAFMTFVIDADQ 331
Db 281 HDKIRKRGVITISGKMTFTSNKSMIEVLVDADPVDSSQKRYPAASAFYTVSLSGCR 340
QY 332 PQLPMTIRPQDGRRRYREASAR 355
Db 341 SLVPVQLVPTEDKKRPEEGKGR 364
RESULT 12
VLDL_HELPY STANDARD; PRT; 174 AA.
AC 005729;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein vldl.
GN VLDL OR HP0891 OR JHP0824.
OS Helicobacter pylori (Campylobacter pylori), and
OC Bacteri; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210, 85963;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49503 / 60190;
RX MEDLINE=97284485; PubMed=9139699;
RA Cao P., Cover T.L.;
RT "High-level genetic diversity in the vapt chromosomal region of
RT Helicobacter pylori."
RL J. Bacteriol. 179:2852-2856 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wattey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter

```

```

RT pylori."
RL Nature 388:539-547 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., deJonge B.L., Carmel G.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Ives C.,
RA Tummino P.J., Caruso A., Tria-Nickelsen M., Mills D.M., Vovis G.F.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180 (1999).
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U94318; AAC45242.1; -.
DR EMBL; AE000599; AAD07940.1; -.
DR EMBL; AF001512; AAD06410.1; -.
DR PIR; C64631; C64631.
DR PIR; T09451; T09451.
DR TIGR; HP0891; -.
DR InterPro; IPR006683; Thioestr_supf.
DR Pfam; PF03061; 4HBT; 1.
KT Hydrolase; Complete proteome.
FT CONFLICT 8
SQ SEQUENCE 174 AA; 19491 MW; 48D2C71C9AD2A534 CRC64;
Query Match
Best Local Similarity 31.4%; Pred. No. 6e-07;
Matches 44; Conservative 28; Mismatches 66; Indels 0; Gaps 0;
QY 45 TEVQMSQVLDPCHTNGRGLSVGQLTKMIDTTCASAEHRHAGCPCTASMDIYFHTIS 104
Db 22 TKLMSYLVVPTTNPNNVHGGELNLDKVAAYCSTRYCAKGVTLSDGVTEKYP 81
QY 105 VGQVNNIKAKYNNRANFNSMEVGIQVASEDLCEKQMNCKALAFVAREITTKVLTQIT 164
Db 82 VGNLLTFLASINIVGNTSCVGIKVLSEDIKTRITHNTNSCYFTWAVENGKPTMPKYE 141
QY 165 PRTEEKXHSVAERBRMR 184
Db 142 PTEVEIRRYEGALKRKEMR 161
RESULT 13
YKHA_BACSU STANDARD; PRT; 179 AA.
AC P49851;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative acyl-CoA thioester hydrolase ykha (EC 3.1.2.-).
GN YKHA OR BSU1030.
OS Bacillus subtilis.
OC Bacteri; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96272254; PubMed=9682784;
RA Lacelle M., Kumano M., Kurita K., Yamane K., Zuber P., Nakano M.M.;
RT "Oxygen-controlled regulation of the flavohemoglobin gene in Bacillus
RT subtilis."
RL J. Bacteriol. 178:3803-3808 (1996).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RT "Sequence of the Bacillus subtilis genome between xlyA and ykor.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berttero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.T., Comerion I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Portolillo S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekituchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
RA Takenchi M., Tamakoshi A., Tanaka T., Terstra P., Togomori K.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambert R., Wedler E., Wedler H., Welterngger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Dandhin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:49-256(1997).
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D78189; BAA11257.1; -.
CC EMBL: AJ002571; CA005683.1; ALT_INIT.
CC EMBL: Z99110; CAB13160.1; ALT_INIT.
CC Subtilast; BG11417; ykha.
CC InterPro: IPR006683; Thioestr_supf.
CC Pfam: PF03061; 4HPT; 1.
CC Hydrophobic protein; Hydrolase; Complete proteome.
CC KW
CC SEQUENCE 179 AA; 20065 MW; 8CB508A96A152B CRC64;

```

```

Db 132 VPETEEMIMHNTAVGRANER 152
RESULT 14
AC48 MOUSE STANDARD; PRT; 439 AA.
ID AC48_MOUSE
AC Q9R0X4; Q9WUJ0; Q9WU28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 48 kDa acyl-CoA thioester hydrolase, mitochondrial precursor
DE (EC 3.1.2.-) (p48) (MC-ACT48) (Protein U8).
GN ACAT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=99315862; PubMed=10383425;
RA Poupon V., Begue B., Gagnon J., Dautry-Varsat A., Cerf-Bennussan N.,
RA Benmehar A.;
RT "Molecular cloning and characterization of MT-ACT48, a novel
RT mitochondrial acyl-CoA thioesterase.";
RL J. Biol. Chem. 274:19188-19194(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AKR;
RA Ishizuka Y., Mochizuki R., Tohdo N.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Active on long chain acyl-CoAs.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC
CC -1- SIMILARITY: Belongs to the acyl coenzyme A hydrolase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ238893; CAB45192.1; -.
CC EMBL: AJ238894; CAB45255.1; -.
CC EMBL: AB028898; BAA79193.1; -.
CC MGD; MGI:1928939; Acate2.
CC InterPro: IPR006683; Thioestr_supf.
CC Pfam; PF03061; 4HPT; 1.
CC KW
CC Hydrolyase; Serine esterase; Mitochondrion; Transit peptide.
CC TRANSIT
CC CHAIN 1 21
CC VARIANT 22 439
CC VARIANT 6 6
CC VARIANT 40 41
CC VARIANT 49 49
CC VARIANT 58 58
CC VARIANT 82 82
CC VARIANT 119 119
CC VARIANT 137 137
CC VARIANT 140 140
CC VARIANT 140 140
CC VARIANT 177 177
CC VARIANT 189 189
CC VARIANT 192 192
CC VARIANT 206 206
CC VARIANT 212 212
CC VARIANT 231 231
CC VARIANT 267 267
CC VARIANT 366 366
CC VARIANT 378 378
CC CONFLICT 219 222
CC SEQUENCE 439 AA; 50560 MW; 309CD950D85ACBD0 CRC64;

```

Query Match 5.5%; Score 175.5; DB 1; Length 179;
 Best Local Similarity 34.8%; Pred. No. 8.3e-07;
 Matches 49; Conservative 22; Mismatches 69; Indels 1; Gaps 1;

Db 105 VGVVNIKAVNBAFNSMVEGIVASSEDCESEKONVCAKALTFVARRITK-VKLKQI 163
 72 QKQVCLSESYVTWVGISMEVFPVKVKEHMTGERELAAISLTFVALDNGKFPVPRV 131
 164 TPRTSEKHEHSAVAERRRMR 184

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:42:01 ; Search time 20 Seconds

(without alignments)
2919.413 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181
Sequence: 1 MIONVGNHLRGLASVFSNR.....TKALPFAKINSRFGYLQDT 607

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:**
1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3181	100.0	630	2	J00351
2	225	7.1	338	2	J05415
3	224.5	7.1	276	2	E98257
4	223.5	7.0	155	2	G72051
5	223.5	7.0	155	2	C86572
6	220	6.9	343	2	J07161
7	218	6.9	343	2	J05416
8	216.5	6.8	157	2	F83749
9	216	6.8	157	2	F83937
10	214.5	6.7	160	2	F71502
11	207.5	6.5	159	2	C81662
12	204	6.4	167	2	A96036
13	204	6.4	318	2	B90399
14	201	6.3	188	2	F82956
15	184	5.8	438	2	H84711
16	180.5	5.7	166	2	E83309
17	179	5.6	160	2	G81079
18	179	5.6	160	2	F81864
19	177	5.6	174	2	T09451
20	177	5.6	174	2	C64631
21	174.5	5.5	365	2	E72725
22	174	5.5	172	2	AB1690
23	173	5.4	172	2	AB1318
24	170.5	5.4	176	2	E89975
25	163.5	5.1	222	2	C75539
26	157	4.9	155	2	H84243
27	151	4.7	147	2	G75629
28	148	4.7	141	2	A83178
29	144.5	4.5	136	2	E69856

30	142.5	4.5	158	2	AF3027	acyl-CoA hydrolase
31	133	4.2	137	2	B81365	probable hydrolyase
32	130.5	4.1	162	2	D82397	acyl-CoA thioester
33	124	3.9	147	2	D84393	acyl-CoA thioester
34	123	3.9	393	2	T25908	hypothetical prote
35	121	3.8	1711	1	A47392	chromodomain-helic
36	118	3.7	445	2	I38027	MLN 64 protein - h
37	114.5	3.6	148	2	E81878	probable acyl-CoA
38	114	3.6	149	2	AF0267	probable acyl-CoA
39	114	3.6	154	2	A64097	conserved hypochet
40	114	3.6	1275	2	T37192	nebulin-related pr
41	112	3.5	461	2	T30073	hypothetical prote
42	111.5	3.5	440	2	T32085	hypothetical prote
43	111	3.5	150	2	C98254	probable acyl-CoA
44	110	3.5	152	2	E87272	hypothetical prote
45	108.5	3.4	130	2	AE2821	acyl-CoA hydrolase

ALIGNMENTS

RESULT 1

T00351
hypothetical protein KIAA0707 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #ext_change 21-Jul-2000
C:Accession: T00351
R:Ichikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A>Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; M01D:98403860; PMID:9734811
A:Accession: T00351
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-630 <ISH>
A:Cross-references: EMBL:AB014607; NID:93327227; PTDN:BA031682.1; PTD:93327228
A:Experimental source: brain
C:Genetics:
A>Note: KIAA0707

Query Match	100.0%;	Score 3181;	DB 2;	Length 630;
Best Local Similarity	100.0%;	Pred. No. 4e-238;		
Matches 607;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MIONVGNHLRRGLASVFSNRTSRKSALRAGNDSAMADGEGYRNPTVEVOMQLVLPCHTNQ	60	
DB	24	MIONVGNHLRRGLASVFSNRTSRKSALRAGNDSAMADGEGYRNPTVEVOMQLVLPCHTNQ	83	
QY	61	RGLSVGOLIKWIDTTACLSAERHAGCPVTASMDIYFEHTISVGVVNIKAKVNAFN	120	
DB	84	RGLSVGOLIKWIDTTACLSAERHAGCPVTASMDIYFEHTISVGVVNIKAKVNAFN	143	
QY	121	SSMEVGIVASSEDSCSEKQNVCKALATFVARREITTKVKQITPREEEKMESHVAER	180	
DB	144	SSMEVGIVASSEDSCSEKQNVCKALATFVARREITTKVKQITPREEEKMESHVAER	203	
QY	181	RRMRVYADTTIKDLANCAIQGDLSDSCSRMTPAEKTRVESVLYLPHPANQGNTPGG	240	
DB	204	RRMRVYADTTIKDLANCAIQGDLSDSCSRMTPAEKTRVESVLYLPHPANQGNTPGG	263	
QY	241	QIAMMENVATTAASRLCRAHPTLKAIMEHFPGSPGVGDLVYKATVNNAPFKHSMIEGV	300	
DB	264	QIAMMENVATTAASRLCRAHPTLKAIMEHFPGSPGVGDLVYKATVNNAPFKHSMIEGV	323	
QY	301	CVSAVYQEAETHRRHINSAPMTFVLDADQPOLLPPIRQPGGERRYREASARKKIRL	360	
DB	324	CVSAVYQEAETHRRHINSAPMTFVLDADQPOLLPPIRQPGGERRYREASARKKIRL	383	
QY	361	DRKYIVSCQTEVPLSPVMPDPSNQVLYSYNNVSLKMLVAKDNWVLSSEISOVLXTLIED	420	
DB	384	DRKYIVSCQTEVPLSPVMPDPSNQVLYSYNNVSLKMLVAKDNWVLSSEISOVLXTLIED	443	
QY	421	DKFLSPHEMVEVAVHVDAAQAFLLISDLRQRPBEMDKHRSVELVQGVDEDDAIYHTVSPALG	480	

```

Db      444 DKFSTHMEMVYHDAQAFLLSDLRORPEMDKHYREVELVQVDEDDALHYHTSPALG 503
QY      481 GHTYKQDFVLIASRRKPCDNDPVYIALRSVTLPTHRETPYRGGTLLCSGFCIMREGDQ 540
Db      504 GHTYKQDFVLIASRRKPCDNDPVYIALRSVTLPTHRETPYRGGTLLCSGFCIMREGDQ 563
QY      541 LTKCCWAVSLTEHVSASGFYSWIGESRSKRRSDGMNKGTLAGHTLTKAIIPAKINSR 600
Db      564 LTKCCWAVSLTEHVSASGFYSWIGESRSKRRSDGMNKGTLAGHTLTKAIIPAKINSR 623
QY      601 FGVLQDT 607
Db      624 FGVLQDT 630

```

RESULT 2

```

JC5415
palmityl-CoA hydrolase (EC 3.1.2.2) brain type - rat
N:Alternate names: long-chain fatty-acyl-CoA hydrolase
C:Species: Rattus norvegicus (Norway rat)
C>Date: 10-Jun-1997 #sequence_revision 12-Sep-1997 #text_change 11-Aug-2003
C:Accession: J05415; PC4318
R:Yamada, Y.; Furuhata, T.; Iida, N.; Watanabe, T.; Hosokawa, M.; Satoh, T.; Someya, A.;
Biochem. Biophys. Res. Commun. 232, 198-203, 1997
A:Title: Molecular cloning and expression of cDNAs encoding rat brain and liver cytosol
A:Reference number: J05415; MUID:97236308; PMID:9125130
A:Accession: J05415
A:Molecule type: mRNA
A:Residues: 1-338 <YAML>
A:Cross-references: DDBJ:D88890; NID:G1944427; PIDN:BAA19626.1; PID:G1944428
A:Experimental source: brain
A:Accession: PC4318
A:Molecule type: Protein
A:Residues: 17-29;213-221;272-300;327-337 <YAML>
A:Experimental source: brain
C:Comment: This enzyme catalyzes the hydrolysis of fatty acyl-CoA thioesters to free fatty
acids and CoA. It is involved in intracellular protein transport, activation of protein kinase and regulation
of cellular metabolism.
C:Keywords: coenzyme A; cholesterol hydrolase
F:314-330/Region: nuclear location signal

```

```

Query Match      7.1%; Score 225; DB 2; Length 338;
Best Local Similarity 22.4%; Pred. No. 1,le-09;
Matches 74; Conservative 66; Mismatches 155; Indels 36; Gaps 7;

```

```

QY      44 PTEVOMQQLVPCHTNORGLSVQQLKWDITTAQSLAERHAG-----CPCTVASMDDI 97
Db      9 PAIQRIMRPDANVAVNGVGGTILKMEAGVILSTRHCNSQNGRCVAAALARVERT 68
QY      98 YFEHTISGVQVNIKAKVNFNSMEVGIQVASEDLCEKQNVCKLAFV--AREI 155
Db      69 DFLSPMCGEVAHVAEITYSKSHVEVQVHLSNLTGKKLTNKAITLVVPLSLKNV 128
QY      156 TKV-----KLKQITRTEEKMEHVAERRRRLVY--ADTIKDLNCAIQDLESRDC 209
Db      129 DKVLEVPPIVLRQOESEKRYEAQLEKEMTKMRKGDIVQILN----- 175
QY      210 SRWPAEKT---RVESEVELVPPHANGTFCGQIWMAMENVATTAAELCRAPHTLK 265
Db      176 ---PENTVSYSSSLIHVGPDCITLHGFVHGVTMKLMDVAGIYAAHCKTNIVTA 231
QY      266 AIEHFHRRGSPQVQDRLVLAIVNNAFHSMEVGCVEAYQOEAFTHRR--INSAPMTFV 324
Db      232 SVDAINHDKIRKQCVITISGRMFTSNKSEIEVLVDADVPVNSQKRYAASAFPTV 291
QY      325 VLDADQPOLPLWTRPQGDGERRRYEASAR 355
Db      292 SLNDEGKPLVPQVLVPEDEDEKRFEEGKGR 322

```

RESULT 3

B98257

```

hypothetical protein AGR_L_2016 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: B98257
R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tume
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B98257
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-276 <XLR>
A:Cross-references: GB:AE007870; PIDN:AAK9583.1; PID:G15159470; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2016
A:Map position: linear chromosome

```

```

Query Match      7.1%; Score 224.5; DB 2; Length 276;
Best Local Similarity 25.2%; Pred. No. 9.4e-10;
Matches 72; Conservative 42; Mismatches 127; Indels 45; Gaps 5;

```

```

QY      48 QMSQVLVPCHTNORGLSVQQLKWDITTAQSLAERHAGCPVYASMDIYFEHTISVQ 107
Db      18 RLIDIVFPGDINHHTLFCGTGLAMRVAFIATFRTPTVYASCRIDFRQPARIGH 77
QY      108 VVNIKAKVNFNSMEVGIQVASEDLCEKQNVCKLAFVAREITKVKLKOITRT 167
Db      78 IVEFTAPVVKAGRSLLTVEVEMVAETIIGRQOHTCTRGIFPMVA-----IPEG 125
QY      168 EEEKMEHVAERRRMRLVADDTKDLNCAIQDLESRDCSMVNPAAKTRVESVELV 227
Db      126 ED-----NASTVLPDLITET-----PPSDATVWEIVF 155
QY      228 PPHANQNTFGQIMAMENVATTAAELCRAPHTLKALIMFHRGSPQVQDRLVLA 287
Db      156 PDQANSAGRMGGAIVMTYAAVFAVSRVCGKLVLASRRIDPARAIBIGEIVQA 215
QY      288 VVNAFKHSMVEVGCVEAYRQAEFTHRRIN--SAPMTFVLVADDP 332
Db      216 VERVGRSSMSIQ--TKLMSENLLTGERHITATGHTVAVADKDRP 259

```

RESULT 4

```

G72051
cytosolic acyl-CoA thioester hydrolase family protein CP0093 [imported] - Chlamydia
N:Alternate names: acyl-CoA thioesterase
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: G72051; A81615
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: G72051
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-155 <AKN>
A:Cross-references: GB:AE001648; GB:AE001363; NID:G4376946; PIDN:AA018793.1; PID:G4376955;
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, F.;
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: A81615
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-155 <REA>
A:Cross-references: GB:AE002172; GB:AE002161; NID:G7189018; PIDN:AA97977.1; PID:G7189027;
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: ycaA; CP0093

```


Db 181 -PEPTVSYSSQSLIHLVGPDSCTLHGFGVGTWKLMDNEVAGIYARHCKTNIIVTASVD 239

Qy 269 MEHFRGSPQVQDRLVTKAIYNNAFKHSMEVGVCEVAYROEAEFTHRRH-INSAPMTFVLDDADDP 327

Db 240 AVNPFDKIRKCGCVTTISGRMFTSNKSMIEVLVDADVPVNSQKRYRAASAFFTYVSLN 299

Qy 328 ADDQPOLPMTIRPOPGDGERRYREASAR 355

Db 300 QEKRLPVQPLVPTEDDEKRRFEQKGR 327

RESULT 8

F83749

acyl-CoA hydrolase BH0798 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: F83749

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: AB3650; MUID:20512582; PMID:11058132

A:Accession: F83749

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-157 <STD>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04517.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0798

Query Match 6.8%; Score 216.5; DB 2; Length 157;

Best Local Similarity 32.0%; Pred. No. 1.7e-09;

Matches 48; Conservative 32; Mismatches 67; Indels 3; Gaps 2;

Qy 214 PAEKTRESEVLVLPHPANHOQNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHR 273

Db 6 PVERSRITQTRLVLPPTNHLGITFGKVLAYIDEIATLTKMANSAYVTASIDSDVDFK 65

Qy 274 GPSQVGDRLVTKAIYNNAFKHSMEVGVCEVAYROEAEFTHRRHNS-APMTFVLDDADDP 332

Db 66 SSATVGDALHEGVTYHGRISME--VYRVHNSNLTGERTLTTESPFLTVAVADESGK 123

Qy 333 QLLPWIRPOPGDGERRYREASARKKIRLDR 362

Db 124 KPVPQVEPQTEEKRLVETAPARKENRKR 153

RESULT 9

F83937

acyl-CoA thioester hydrolase BH2302 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: F83937

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: AB3650; MUID:20512582; PMID:11058132

A:Accession: F83937

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-162 <STD>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06021.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2302

Qy 274 GPSQVGDRLVTKAIYNNAFKHSMEVGVCEVAYROEAEFTHRRHIN-SAPMTFVLDDADDP 332

Db 66 SPIRTGEALICLEGITVYTRKTSMEVFAVVEA--EDLTGERRLTATSYLTFVALNPDGK 123

Qy 333 QLLPWIRPOPGDGERRYREASARKKIRLDRK 363

Db 124 AEIPPVPTTEDEKKNHYGAKEREYRIRRR 154

RESULT 10

F71502

probable acyl-coa thioesterase - *Chlamydia trachomatis* (serotype D, strain W3/Cx)

C:Species: *Chlamydia trachomatis*

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999

C:Accession: F71502

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia tract*

A:Reference number: A71570; MUID:99000809; PMID:9784136

A:Accession: F71502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-160 <ARN>

A:Cross-references: GB:AE001324; GB:AE001273; NID:g3328966; PIDN:AAC68137.1; PID:g332897;

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: yC1A

Query Match 6.7%; Score 214.5; DB 2; Length 160;

Best Local Similarity 34.1%; Pred. No. 2.6e-09;

Matches 47; Conservative 30; Mismatches 54; Indels 7; Gaps 2;

Qy 226 VLPPHANHOQNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHRGSPQVQDRLVTK 285

Db 25 IFPNDLNTNNTIFGGLMSLDRALVVAERHCESICVTAFVDAMRFYAPAYMGENTLIC 84

Qy 286 AIYNNAFKHSMEVGVCEA--YRQEAETHRRHNSAPMTFVLDDADDPQLPWIRPO 342

Db 85 ASVRSKMTSLVGVKVAENITKOE---HRHITSALTFTVAVDKNSPVEBELTES 140

Qy 343 GDGRRRYREASARKKIRL 360

Db 141 QEBIRRFREADQRRALRL 158

RESULT 11

C81662

cytosolic acyl-CoA thioester hydrolase family protein TC0822 [imported] - *Chlamydia muric*

C:Species: *Chlamydia muridarum*, *Chlamydia trachomatis* MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: C81662

R:Read, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, E

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.

A:Reference number: AB1500; MUID:20150255; PMID:11064935

A:Accession: C81662

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-159 <TEN>

A:Cross-references: GB:AE002348; GB:AE002160; NID:g7190839; PIDN:AAF39624.1; PID:g7190850

A:Experimental source: strain Ni9g (MoPn)

C:Genetics:

A:Gene: TC0822

Query Match 6.5%; Score 207.5; DB 2; Length 159;

Best Local Similarity 33.3%; Pred. No. 8.8e-09;

Matches 46; Conservative 30; Mismatches 55; Indels 7; Gaps 2;

Qy 226 VLPPHANHOQNTFGQIMAMMENAVTIAASRLCRAHPTLKAIEMFHRGSPQVQDRLVTK 285

Db 24 IFPNDLNTNNTIFGGLMSLDRALVVAERHCESICVTALVDVRFYAPAYMGENTLIC 83

H84711

hypothetical protein At2g30720 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence revision 02-Feb-2001

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 11-Aug-2003
C/Accession: H84711

C; ACCESSION: H64/11
R; Lin, X.; Kaul, S.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, C.

Nature 402, 761-768, 1999

A; Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A: Reference number: A94430. MIMD-30083487. PMID-10617197

A:Accession: H84711
A:reference number: A84420; MUID:20083487; PMID:10617197

```

//Accession: 103721
A/Status: preliminary

```

A;Molecule type: DN

A;Residues: 1-438 <

A; Cross-references: GB:
C: Genetics;

C;Genetics:
A:Gene: AT2g30720

```

r;Scene: 00000000
A;Map position: 2

```

C; Superfamily: ac

1

Query Match	5.8%; SC
Post-Test 1 Similarity	0.0%

Best Local Similarity 25.0%; Pred. No. 2.5e-06;
Matches 96; Conservative 51; Mismatched 171; Indel 1

Matches 96; Conservative 51; Mismatches 171; Indels 66; Gaps 13;

39 EGYRN-----PTEVQMSOLVLP-----HTNORGELSVGOLTKWIDTTACLSAERHAG-- 86

Category	Sub-category	Value
X	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
Y	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1
Z	1	1
	2	1
	3	1
	4	1
	5	1
	6	1
	7	1
	8	1
	9	1
	10	1

Db 77 EARRNFEIPTGDASQSKLTAKSPSRSRTPNPWNEIRTGKLVEDLDALAGTISFKHCGGD 136

1. 4. 2019

QY 87-----CPCVTASMDDIYFEHTISVGQVNIKAKVNRAFNSSMEVGIQVAS-EDLCSEKQW 140

Db 137 SSARSMILVTASVDRIIMKPIRVDTLSIVGAVTWVGSSMEMOLOYOTOTNNSSSES 196

10 / DONORSHIP AND VOLUNTARY CONTRIBUTIONS TO THE UNITED STATES 1980-1989

QY 141 NVKALATFVARREIT--KVKLKQITPRTEEKMEHSVAERRMR-----L 185

100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611

Db 197 VALEANFTFVARDAQTGKSAPINQVVPETEHEKFLWKEAERNKLRQKRAQCKEHEKL 256

OV 186 VYADTIKDLANCAIOGDI.ESRDCSRMVPAEKTRVESVEI.VT.PBHANHOGNTEGGOTMAW 245

100 V IAD I I D D E A N C A I Q D D E S K D C S R M V F A E K I R V E S V E L V L P P H A N H Q G N I F G G Q I M A W 245

Db 257 KDLERINELLAEGRVFLDMPALADRNILIKDTSHENSLICQPQRNIHGRIFGGFLMRK 316

[illegible]

QY 246 MENVATIAASRLCRAHPTLKAIEMFHRGSPSQVDRLVLKA-IVNNAFKHSME--VGVCV 302

Search completed: August 12, 2004, 15:45:52
Job time : 22 secs

Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2004, 15:40:21 ; Search time 58 Seconds

(without alignments)
2957.006 Million cell updates/sec

Title: US-10-055-624B-2

Perfect score: 3181

Sequence: 1 MTQNVGNHURRGIAVFSNR.....TLKALPVAKINSRFGYLODT 607

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: Geneseqp29tano4:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	89.2	594	4	AAB95601 Human pro
2	970	30.5	246	4	AAU18380 Human end
3	956	30.1	207	4	AAU20561 Human sec
4	956	30.1	207	4	AAU18487 Human end
5	413	13.0	119	3	AAB58891 Breast an
6	255	8.0	339	4	AAB76624 Cotyrbac
7	255	8.0	339	4	AAG92590 C glutam
8	225	7.1	338	7	ADBS6432 Rat Prote
9	224.5	7.0	371	5	ABP69844 Human pol
10	223.5	6.9	155	6	AAV35275 Chlamydia
11	223.5	6.9	339	2	AAW74896 Protein e
12	220	6.9	339	2	ABG95348 Human nov
13	220	6.9	339	2	ABG95348 Region of
14	220	6.9	339	2	ABG95348 Human pro
15	219	6.9	339	2	ABG95348 Human pro
16	218.5	6.9	339	2	ABG95348 Human pro
17	207.5	6.5	159	6	AAU73682 Amino aci
18	206.5	6.5	375	3	AAV71115 Protein e
19	203	6.4	168	6	ABU17829 Human Hyd
20	191.5	6.0	172	6	ADA34404 Actineta
21	188	5.9	389	4	ABG13806 Novel hum
22	186.5	5.9	177	5	ABP39851 Staphyloc
23	186	5.8	356	6	ABU17697 Protein e
24	186	5.8	370	6	ABU41179 Human DIT
25	185	5.8	176	6	ABU42699 Protein e

26	182	5.7	455	5	AAU77063
27	181.5	5.7	163	6	ABU35445
28	181	5.7	160	6	ABP79796
29	179	5.6	389	4	ABBS8313
30	177	5.6	174	6	ABU09005
31	175.5	5.5	404	4	ABG06274
32	173	5.4	172	5	ABBA8981
33	173	5.4	170	6	ABU32543
34	171.5	5.4	177	6	ABU17573
35	171.5	5.4	177	6	ABU34445
36	170.5	5.4	176	6	ABU42462
37	170.5	5.4	181	6	ABM71499
38	170.5	5.4	472	4	ABBS8527
39	166	5.2	178	6	ABU28987
40	148	4.7	385	4	ABBS8316
41	145	4.6	439	2	AAW74418
42	142	4.5	448	3	AAV71118
43	140.5	4.4	438	5	AAU77062
44	140	4.4	137	6	ABU22724
45	139.5	4.4	131	6	ABU21871

ALIGNMENTS

RESULT 1
AAB95601
ID AAB95601 standard; protein; 594 AA.

AC AAB95601;
DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:18290.

Human, primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-MAY-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto T;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 18290; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 cDNAs, and (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH1632 to AAH1632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention

SQ Sequence 594 AA:

Query Match 89.2%; Score 2838; DB 4; Length 594;

Best Local Similarity 100.0%; Pred. No. 6.9e-278;

Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIVGVGHLRGLASVPSNRKTSKSAIRAGNDSAMADSEGRNPRFVMSQLVPCHTNQ 60
DB 1 MIVGVGHLRGLASVPSNRKTSKSAIRAGNDSAMADSEGRNPRFVMSQLVPCHTNQ 60
QY 61 RGEISVGLLKMIDTTACTAERHAGCCPVASMDIYFEHTISVGQVNIKAKVRAFN 120
DB 61 RGEISVGLLKMIDTTACTAERHAGCCPVASMDIYFEHTISVGQVNIKAKVRAFN 120
QY 121 SSMVEGIVASEDCISEKQNVKALATFVARREITVKYKQITPRTEERMEHSAER 180
DB 121 SSMVEGIVASEDCISEKQNVKALATFVARREITVKYKQITPRTEERMEHSAER 180
QY 181 RRMVLVADTTKDLLANCAIGDLESPDCSRNPAEKTRVSEVLVPPHANHQNFTGG 240
DB 181 RRMVLVADTTKDLLANCAIGDLESPDCSRNPAEKTRVSEVLVPPHANHQNFTGG 240
QY 241 QIMAMMENVAITIASRLCRAPPTKAIEMFHRGSPQVGDVLKAIYNNAFKISMVEGV 300
DB 241 QIMAMMENVAITIASRLCRAPPTKAIEMFHRGSPQVGDVLKAIYNNAFKISMVEGV 300
QY 301 CVEAYRQEAETRRHINSAPFTFVLADADQPOLLPWIRPQGGERRRYSARKKIRL 360
DB 301 CVEAYRQEAETRRHINSAPFTFVLADADQPOLLPWIRPQGGERRRYSARKKIRL 360
QY 361 DRKIVSCQKTEVPLSVWDPNSNOYVLSYNNVSSIKMLVAQDNWLSSEISQVRLYTLED 420
DB 361 DRKIVSCQKTEVPLSVWDPNSNOYVLSYNNVSSIKMLVAQDNWLSSEISQVRLYTLED 420
QY 421 DKFLSFHMEVAVHADAQAFLLSDLRPEWMDKHYRSEVLVQGVDEBDATYHTSPALG 480
DB 421 DKFLSFHMEVAVHADAQAFLLSDLRPEWMDKHYRSEVLVQGVDEBDATYHTSPALG 480
QY 481 GHTYRQDFVILIASRRKPCDNGDPYIALRSVTLPTHRETPYRGETYCSGFCIMREGDQ 540
DB 481 GHTYRQDFVILIASRRKPCDNGDPYIALRSVTLPTHRETPYRGETYCSGFCIMREGDQ 540
QY 541 LTK 543
DB 541 LTK 543

RESULT 2
AAU18380 standard; protein, 246 AA.

AAU18380;

21-NOV-2001 (first entry)

Human endocrine polypeptide SEQ ID No 335.

Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;

KW dog; chicken; sheep; immunosuppressive; antiarthritis; vasotropic;
KW antineoplastic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nocotropic; antibacterial; virinide; fungicide; cancer;
KW ophthalmological; vinnaric; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility.

Homo sapiens.

MO20015364-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001308.

PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0226279P.
PR 18-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.

PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM,
XX
XX WPI; 2001-451931/48.
XX N-PSDB; AAS33270.
XX
XX New nucleic acids and polypeptides, useful for diagnosing, preventing or
XX treating medical conditions.
XX
XX Claim 11; SEQ ID NO 553; 753pp; English.
XX
XX The invention relates to novel isolated nucleic acid molecules (I)
XX encoding human secreted proteins (II). (I) and (II) are used to prevent,
XX treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
XX goats, horses, cats, dogs, chickens or sheep. (II) and (II) may be used in
XX the prevention, treatment and diagnosis of diseases associated with
XX inappropriate expression of secreted proteins. (I) and complementary
XX sequences may also be used as DNA probes in diagnostic assays (e.g.
XX polymerase chain reactions (PCR)) to detect and quantitate the presence
XX of similar nucleic acid sequences in samples, and so which patients may
XX be in need of restorative therapy. (II) may also be used as antigens in
XX the production of antibodies and in assays to identify modulators
XX (agonists and antagonists) of the expression and activity of the secreted
XX proteins. The anti-(II) antibodies and antagonists may also be used to
XX down regulate expression and activity of (II). The anti-(II) antibodies
XX may also be used as diagnostic agents for detecting the presence of (II)
XX in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
XX disorders include for example: immune/autoimmune diseases (e.g. HIV
XX human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
XX and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
XX melanomas, neoplasms of the breast or liver, Sezary syndrome and
XX Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
XX Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
XX cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
XX thrombosis), infections caused by bacteria, viruses and fungi and ocular
XX disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
XX and antibodies can also be used to promote wound healing, maintain organs
XX before transplantation, and support cell culture of primary tissues.

Query Match 30.1%; Score 956; DB 4; Length 207;
Best Local Similarity 95.3%; Pred. No. 6, 9e-88;
Matches 181; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 360 LBRKYVSCQKQTEVPLSVDPSPNOVYLSYNNVSLKMLVAKDNWVLSSEISQVRLTYLLE 419
DB 12 LPRKYVSCQKQTEVPLSVDPSPNOVYLSYNNVSLKMLVAKDNWVLSSEISQVRLTYLLE 71
QY 420 DDKEFSFHEMNVHVAQAQFLLISDLQRPEWDKHYRSELYVOQVEDDAIYHVSPAL 419
DB 72 DDKEFSFHEMNVHVAQAQFLLISDLQRPEWDKHYRSELYVOQVEDDAIYHVSPAL 131
QY 480 GGHTRQDFVYLLASRRKPCDNDPVYALASVTLPTRETPRYRGGTLCGFCMLMREGD 539
DB 132 GGHTRQDFVYLLASRRKPCDNDPVYALASVTLPTRETPRYRGGTLCGFCMLMREGD 191
QY 540 QLTXXCMVRY 549
DB 192 QLTXXCMVRY 201

RESULT 4

AAU18487
ID AAU18487 standard; protein; 207 AA.
XX
XX AAU18487;
AC
XX
DT 21-NOV-2001 (first entry)
XX
XX Human endocrine polypeptide SEQ ID No 442.
XX
XX Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat;
XX dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
XX cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
XX ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
XX hyperproliferative disorder; breast; liver; cardiovascular disorder;
XX cerebrovascular disorder; nervous system disorder; bacterial infection;
XX fungal infection; viral infection; ocular disorder; endocrine disorder;
XX gastrointestinal disorder; renal disorder; respiratory disorder;
XX wound healing; skin aging; organ transplantation; food preservative;
XX tissue regeneration; anti-fertility.
XX
XX Homo sapiens.
XX
XX WO200155364-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001308.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190766P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 06-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.

[illegible]

Db 192 QLTWACSRL 201

RESULT 5
AAB58891

ID AAB58891 standard; protein; 119 AA.

AC AAB58891;

XX

XX

DT 27-MAR-2001 (first entry)

XX

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 599.

XX

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.

XX

OS Homo sapiens.

XX

XX WO200055173-A1.

XX

XX 21-SEP-2000.

XX

XX 08-MAR-2000; 2000WO-US005881.

XX

XX 12-MAR-1999; 99US-0124270P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WFI; 2000-611515/58.

XX

DR N-PSDB; AAF21794.

XX

PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention, treatment
PT and diagnosis of cancer, immune disorders, cardiovascular disorders and
PT neurological diseases.

XX

PS Claim 11; Page 1036; 12999p; English.

XX

CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic, immunosuppressive, neotropic;
CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
CC antiinflammatory; antitumor; anticonvulsant; antibacterial;
CC antifungal; antiparasitic and cardiac activity. The polynucleotide and
CC protein sequences are used in the diagnosis of cancer, particularly
CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists
CC and antagonists may also be used in the diagnosis, prevention and treatment
CC of immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC cardiovascular disorders such as myocardial ischemias; wound healing;
CC neurological diseases such as cerebral anoxia and epilepsy; and
CC infectious diseases

XX

XX

SQ Sequence 119 AA;

Query Match 13.0%; Score 413; DB 3; Length 119;
Best Local Similarity 96.4%; Pred. No. 3e-33;
Matches 81; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 383 NOVYLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLLEDDKFLSPHEMNVVVDAAQAFLL 442

Db 4 DQVYLSYNNVSSIKMLVAKDNWVLSSEISQVRLTYLLEDDKFLSPHEMNVVVDAAQAFLL 63

QY 443 LSDLRQRPENMDKHYRSVELVQOVD 466

Db 64 LSDLRQRPENMDKHYRSVELVQOVD 87

RESULT 6
AAB76624

ID AAB76624 standard; protein; 339 AA.

AC AAB76624;

XX

XX

DT 11-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum MCT protein SEQ ID NO:230.

XX

XX Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.

XX

XX

XX Corynebacterium glutamicum.

XX

XX WO200100805-A2.

XX

XX 04-JAN-2001.

XX

XX 23-JUN-2000; 2000WO-IB000926.

XX

XX 25-JUN-1999; 99US-0141031P.

XX

XX 08-JUL-1999; 99DE-01031454.

XX

XX 08-JUL-1999; 99DE-01031478.

XX

XX 08-JUL-1999; 99DE-01031563.

XX

XX 09-JUL-1999; 99DE-01032122.

XX

XX 09-JUL-1999; 99DE-01032124.

XX

XX 09-JUL-1999; 99DE-01032125.

XX

XX 09-JUL-1999; 99DE-01032128.

XX

XX 09-JUL-1999; 99DE-01032180.

XX

XX 09-JUL-1999; 99DE-01032182.

XX

XX 09-JUL-1999; 99DE-01032191.

XX

XX 09-JUL-1999; 99DE-01032209.

XX

XX 09-JUL-1999; 99DE-01032212.

XX

XX 09-JUL-1999; 99DE-01032227.

XX

XX 09-JUL-1999; 99DE-01032228.

XX

XX 09-JUL-1999; 99DE-01032229.

XX

XX 14-JUL-1999; 99DE-01032230.

XX

XX 14-JUL-1999; 99DE-01032927.

XX

XX 14-JUL-1999; 99DE-01033005.

XX

XX 14-JUL-1999; 99DE-01033006.

XX

XX 27-AUG-1999; 99DE-01040764.

XX

XX 27-AUG-1999; 99DE-01040765.

XX

XX 27-AUG-1999; 99DE-01040766.

XX

XX 27-AUG-1999; 99DE-01040810.

XX

XX 27-AUG-1999; 99DE-01040831.

XX

XX 27-AUG-1999; 99DE-01040832.

XX

XX 27-AUG-1999; 99DE-01040833.

XX

XX 31-AUG-1999; 99DE-01041378.

XX

XX 31-AUG-1999; 99DE-01041379.

XX

XX 31-AUG-1999; 99DE-01041395.

XX

XX 03-SEP-1999; 99DE-01042077.

XX

XX 03-SEP-1999; 99DE-01042078.

XX

XX 03-SEP-1999; 99DE-01042079.

XX

XX 03-SEP-1999; 99DE-01042088.

XX

XX (BADI) BASF AG.

XX

XX Pompejus M, Kroegeer B, Schroeder H, Zeider O, Heberhauer G;

DR WPI: 2001-071486/08.
DR N-PSDB; AAF67857.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
XX transformation.
XX
PS Claim 20; Page 495-496; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAF676510 to
CC AAF676447. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention
XX
SQ Sequence 339 AA;
Query Match 8.0%; Score 255; DB 4; Length 339;
Best Local Similarity 25.7%; Pred. No. 1.8e-16;
Matches 89; Conservative 56; Mismatches 171; Indels 30; Gaps 10;
QY 41 YRNPLEVQMSQVLPCHTNQSGELSV--GQLKMTDTTACLSAERHAGCPVTASMDIY 98
DB 14 HRSP-EVTLRMAAPTDVLMAGSHGVGGVLEMDKAAYACATQWSGTCVTAYVGHIIH 72
QY 99 FEHTISVQGVNIAKAVNAFNSMEVGIQVASEDLCEKQMNVCALATVAREIT-- 156
DB 73 FTRPIPSGHWVEVSRIMTGRSSMHIYNEVLASDPDGNTRACDCLIVIAKDTATGR 132
QY 157 KVKLKQITPRTEERKMEHSAERRRMLVYADITIKDLANCAIOGDIESDCCSRMPAE 216
DB 133 ATPVPSFTPKNEEQRVLEANSRIGLRK-----ALAEWEKQTYNG--PSE 177
QY 217 KTRVESVELVPRPHANQGNFTGGQIMAMMENVAITIASRLCRAHPTLKAIEHFRRGPS 276
DB 178 APRILITFLAKPTDINMGKHYHGGTAMEMIDEAGACTMWSGNTTAVVAGIRFYQPI 237
QY 277 QVGRDLVYKAIVNNAFKHMEVGVCV--EAYROBAETHRR-HINSAPMTFVLADADQP 332
DB 238 QIGDLIEVDAMRTDKRSQMOSIHVRAGDAHRGAELETAIH--ATVTYIGIDVDGEP 294
QY 333 QLLFWIRPQPGDGERRYREASARKKIRLDRKYI--VSCQTEVPLSV 377
DB 295 LPAPQFVPRTPEDIQLAEHANILRLDRAD--YTPMPLFQRRVPLQI 338
RESULT 7
AA92590
ID AA92590 standard; protein, 339 AA.
XX
XX AAG92590;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum protein fragment SEQ ID NO: 6344.
DE
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
PN
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-00127688.
PF
XX

PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR
DR N-PSDB; AAF67809.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 17; SEQ ID NO 6344; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Coryneform bacterium, and identifying a homologue of a gene derived from
CC coryneform bacterium. Coryneform bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 339 AA;
Query Match 8.0%; Score 255; DB 4; Length 339;
Best Local Similarity 25.7%; Pred. No. 1.8e-16;
Matches 89; Conservative 56; Mismatches 171; Indels 30; Gaps 10;
QY 41 YRNPLEVQMSQVLPCHTNQSGELSV--GQLKMTDTTACLSAERHAGCPVTASMDIY 98
DB 14 HRSP-EVTLRMAAPTDVLMAGSHGVGGVLEMDKAAYACATQWSGTCVTAYVGHIIH 72
QY 99 FEHTISVQGVNIAKAVNAFNSMEVGIQVASEDLCEKQMNVCALATVAREIT-- 156
DB 73 FTRPIPSGHWVEVSRIMTGRSSMHIYNEVLASDPDGNTRACDCLIVIAKDTATGR 132
QY 157 KVKLKQITPRTEERKMEHSAERRRMLVYADITIKDLANCAIOGDIESDCCSRMPAE 216
DB 133 ATPVPSFTPKNEEQRVLEANSRIGLRK-----ALAEWEKQTYNG--PSE 177
QY 217 KTRVESVELVPRPHANQGNFTGGQIMAMMENVAITIASRLCRAHPTLKAIEHFRRGPS 276
DB 178 APRILITFLAKPTDINMGKHYHGGTAMEMIDEAGACTMWSGNTTAVVAGIRFYQPI 237
QY 277 QVGRDLVYKAIVNNAFKHMEVGVCV--EAYROBAETHRR-HINSAPMTFVLADADQP 332
DB 238 QIGDLIEVDAMRTDKRSQMOSIHVRAGDAHRGAELETAIH--ATVTYIGIDVDGEP 294
QY 333 QLLFWIRPQPGDGERRYREASARKKIRLDRKYI--VSCQTEVPLSV 377
DB 295 LPAPQFVPRTPEDIQLAEHANILRLDRAD--YTPMPLFQRRVPLQI 338
RESULT 8
ADE56432
ID ADE56432 standard; protein, 338 AA.
XX
XX ADE56432;
AC
XX 29-JAN-2004 (first entry)
DT
XX Rat Protein BAA19626, SEQ ID NO 2285.
DE
XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KM

CC directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 371 AA;

Query Match 7.1%; Score 224.5; DB 5; Length 371;
 Best Local Similarity 22.4%; Pred. No. 2.5e-13;
 Matches 83; Conservative 66; Mismatches 174; Indels 47; Gaps 8;

QY 9 LRRGLAVSFNSRTSRKSKALRAGNDSAMA---DDEGYNPTEVQMSQVLVPCHTNNGEL 64
 DB 10 LRRGLPPTC-----ALLQPPASAAAAPMSGPDVETPSAIQICRLMRDDANVAGNV 62
 QY 65 SVGQLMKWIDTACISERHAG-----CPCVTASMDIYEHTISVGOVNIKAKYVRA 118
 DB 63 HGGTILKMIERAGAIISTRHCNSONGERCVALARVENTDLSMCGICEVAHVABEITYT 122
 QY 119 FNSMEVGIQVASEDLCEKQWNCALATFVARREITVKLKQITP-----RTBEK 172
 DB 123 SKHSVEVQVNVMSHNLTGAKKLTNKATLWYVPLSKVDKYLEVPVYSGROQEEGR 182
 QY 173 EHSVAERRRRLVY--ADTIKDLNCAIQDLESRCSPVPAEKT---RVESVELY 226
 DB 183 KRYPAQKLERMETKRNQDIVQPVLN-----DEPNTVSYSQSSLIHLV 225
 QY 227 LPPHANHQNTFGQIMAMMENAVTTIASRLCRAPHTLKAIEHFHFRGSPQVGRVLVKA 286
 DB 226 GPSCTLHGFPHGVTKMLDEBAGIVAKCKINIVASVDALNFHKIKKGCYITISG 285
 QY 287 IVNNAFKHSMVEVGVCEAYRQEAETHRRH-INSAPMTFVLADADQPOLLPWIRPQDG 345
 DB 286 RMTFTSNKSEIEVLVDADPVVDSSQKRYRASAFTVYLSQGRSLPVPQLVETDE 345
 QY 346 ERRYREASAR 355
 DB 346 KKRFEKGKR 355

RESULT 10

AA35275
 ID AAY35275 standard; protein, 155 AA.

XX AAY35275;
 AC
 XX 17-OCT-2003 (revised)
 DT 13-SEP-1999 (first entry)
 XX
 DE Chlamydia pneumoniae transmembrane protein sequence.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;
 KW neutralising epitope.

XX Chlamydia pneumoniae.

XX WO9927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB001890.

XX 21-NOV-1997; 97FR-00014673.

XX 04-NOV-1998; 98US-0107078P.

XX (GENSET) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae.

XX Page 1101; Disclosure; 1912pp; English.

CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae. C.
 CC pneumoniae causes respiratory disease such as pneumonia, bronchitis,
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The
 CC polypeptides encoded by the open reading frames of the C. pneumoniae
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also
 CC be used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17
 CC -OCT-2003 to standardise OS field)

XX Sequence 155 AA;

Query Match 7.0%; Score 223.5; DB 2; Length 155;
 Best Local Similarity 35.5%; Pred. No. 7.5e-14;
 Matches 50; Conservative 29; Mismatches 55; Indels 7; Gaps 2;
 QY 226 VLPPHANHQNTFGQIMAMMENAVTTIASRLCRAPHTLKAIEHFHFRGSPQVGRVLVKA 285
 DB 19 IFPNDIANNNTVFGGLMSLDRLALVVAERHTESVCVTAFDALRFYAPAMGENLICK 78
 QY 286 AIVNNAFKHSMVEVGVCEA---YRQEAETHRRHINSAPMTFVLADADQPOLLPWIRPQ 342
 DB 79 AAVNRTRTSLEVGKVAENIYKOE---RRHTSAVFTVAVVEDNQPIPVHIVBET 134
 QY 343 GDGERRYREASARKKIRLDRK 363
 DB 135 PEKRRYNEADRRRQARLEK 155

RESULT 11

ABU26786
 ID ABU26786 standard; protein, 155 AA.

XX ABU26786;

XX 23-OCT-2003 (revised)

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #12313.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Chlamydia pneumoniae.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PsDB; ACA30656.

XX New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 54710; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of

the 6212 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp://wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to standardize OS field)

OY 260 AHPTIKATEMHERGSPQVGRILVUKATVNNAFKHSMEVGYEAYRQEAETHRRH-INS 318
Db 226 TNYVTASVDALNFHDKIKKCVITISGRMTFTSNKSMELIYLVDDPYVDSQKKRYAAS 285
OY 319 AFMTFVVLDADDQFOLLPWIRPQDGBRRRYEASAR 355
Db 286 AFFTVVSLSQEGRSLLPVQVLVETEDKKRFEKGR 322

RESULT 14
ABO34542
ID ABO34542 standard; protein; 339 AA.
XX AC ABO34542;
XX DT 22-SEP-2003 (first entry)
XX DE Region of human secreted protein encoded by cDNA sequence #169.
XX
KM Human; secreted protein; hyperproliferative disorder; leukaemia;
KM breast cancer; wound; reproductive disorder; blood related disorder;
KM haemophilia; thrombocytopaenia; immunodeficiency; lymph hypoplasia;
KM Wiskott-Aldrich syndrome; autoimmune disorder; multiple sclerosis;
KM graft-versus-host disease; Hashimoto's thyroiditis; allergy; asthma;
KM viral infection; bacterial infection; fungal infection; AIDS; sepsis;
KM renal disorder; kidney failure; cardiovascular disorder; cytostatic;
KM angina pectoris; cerebral ischemia; congenital heart defect;
KM respiratory disorder; neurological disorder; Alzheimer's disease;
KM Parkinson's disease; inflammation; Crohn's disease; vulnery;
KM immunosuppressive; antibacterial; haemostatic; thrombolytic;
KM anticoagulant; neuroprotective; chymotrypsin; antiallergic;
KM antiasthmatic; virucide; fungicide; anti-HIV; nephrotoxic; anti-anginal;
KM cerebroprotective; cardiac; nootropic; antiparkinsonian;
KM antiinflammatory.
XX
XX Homo sapiens.
XX OS
XX US2003049618-A1.
XX PD 13-MAR-2003.
XX PF 16-MAR-2001; 2001US-00809391.
XX
PR 07-MAR-1997; 97US-0038621P.
PR 07-MAR-1997; 97US-0040162P.
PR 07-MAR-1997; 97US-0040163P.
PR 07-MAR-1997; 97US-0040333P.
PR 07-MAR-1997; 97US-0040334P.
PR 07-MAR-1997; 97US-0040336P.
PR 11-APR-1997; 97US-0043311P.
PR 11-APR-1997; 97US-0043312P.
PR 11-APR-1997; 97US-0043313P.
PR 11-APR-1997; 97US-0043314P.
PR 11-APR-1997; 97US-0043315P.
PR 11-APR-1997; 97US-0043568P.
PR 11-APR-1997; 97US-0043569P.
PR 11-APR-1997; 97US-0043576P.
PR 11-APR-1997; 97US-0043578P.
PR 11-APR-1997; 97US-0043580P.
PR 11-APR-1997; 97US-0043669P.
PR 11-APR-1997; 97US-0043670P.
PR 11-APR-1997; 97US-0043671P.
PR 11-APR-1997; 97US-0043672P.
PR 11-APR-1997; 97US-0043674P.
PR 23-MAY-1997; 97US-0047492P.
PR 23-MAY-1997; 97US-0047500P.
PR 23-MAY-1997; 97US-0047501P.
PR 23-MAY-1997; 97US-0047502P.
PR 23-MAY-1997; 97US-0047503P.
PR 23-MAY-1997; 97US-0047581P.
PR 23-MAY-1997; 97US-0047582P.

PR 23-MAY-1997; 97US-0047583P.
PR 23-MAY-1997; 97US-0047584P.
PR 23-MAY-1997; 97US-0047585P.
PR 23-MAY-1997; 97US-0047586P.
PR 23-MAY-1997; 97US-0047587P.
PR 23-MAY-1997; 97US-0047588P.
PR 23-MAY-1997; 97US-0047589P.
PR 23-MAY-1997; 97US-0047590P.
PR 23-MAY-1997; 97US-0047592P.
PR 23-MAY-1997; 97US-0047593P.
PR 23-MAY-1997; 97US-0047594P.
PR 23-MAY-1997; 97US-0047595P.
PR 23-MAY-1997; 97US-0047596P.
PR 23-MAY-1997; 97US-0047597P.
PR 23-MAY-1997; 97US-0047598P.
PR 23-MAY-1997; 97US-0047599P.
PR 23-MAY-1997; 97US-0047600P.
PR 23-MAY-1997; 97US-0047601P.
PR 23-MAY-1997; 97US-0047612P.
PR 23-MAY-1997; 97US-0047613P.
PR 23-MAY-1997; 97US-0047614P.
PR 23-MAY-1997; 97US-0047615P.
PR 23-MAY-1997; 97US-0047617P.
PR 23-MAY-1997; 97US-0047618P.
PR 23-MAY-1997; 97US-0047632P.
PR 23-MAY-1997; 97US-0047633P.
PR 06-JUN-1997; 97US-0048964P.
PR 06-JUN-1997; 97US-0048974P.
PR 13-JUN-1997; 97US-0049610P.
PR 08-JUL-1997; 97US-0051926P.
PR 16-JUL-1997; 97US-0052874P.
PR 18-AUG-1997; 97US-0055724P.
PR 22-AUG-1997; 97US-0056630P.
PR 22-AUG-1997; 97US-0056631P.
PR 22-AUG-1997; 97US-0056632P.
PR 22-AUG-1997; 97US-0056636P.
PR 22-AUG-1997; 97US-0056637P.
PR 22-AUG-1997; 97US-0056662P.
PR 22-AUG-1997; 97US-0056664P.
PR 22-AUG-1997; 97US-0056845P.
PR 22-AUG-1997; 97US-0056862P.
PR 22-AUG-1997; 97US-0056864P.
PR 22-AUG-1997; 97US-0056872P.
PR 22-AUG-1997; 97US-0056874P.
PR 22-AUG-1997; 97US-0056875P.
PR 22-AUG-1997; 97US-0056877P.
PR 22-AUG-1997; 97US-0056879P.
PR 22-AUG-1997; 97US-0056880P.
PR 22-AUG-1997; 97US-0056881P.
PR 22-AUG-1997; 97US-0056882P.
PR 22-AUG-1997; 97US-0056884P.
PR 22-AUG-1997; 97US-0056886P.
PR 22-AUG-1997; 97US-0056887P.
PR 22-AUG-1997; 97US-0056888P.
PR 22-AUG-1997; 97US-0056889P.
PR 22-AUG-1997; 97US-0056892P.
PR 22-AUG-1997; 97US-0056893P.
PR 22-AUG-1997; 97US-0056894P.
PR 22-AUG-1997; 97US-0056903P.
PR 22-AUG-1997; 97US-0056908P.
PR 22-AUG-1997; 97US-0056909P.
PR 22-AUG-1997; 97US-0056910P.
PR 22-AUG-1997; 97US-0056911P.
PR 05-SEP-1997; 97US-0057650P.
PR 05-SEP-1997; 97US-0057669P.
PR 05-SEP-1997; 97US-0057761P.
PR 12-SEP-1997; 97US-0058785P.
PR 09-OCT-1997; 97US-0061660P.
PR 06-MAR-1998; 98WO-US004493.
PR 08-SEP-1998; 98US-00149476.
PR 17-MAR-2000; 2000US-0190068P.

XX (RUBEN S. M.
PA (ROSE/ ROSEN C. A.
PA (SOPP/ SOPPET D. R.
PA (CART/ CARTER K. C.
PA (BEDN/ BEDNARIK D. P.
PA (ENDR/ ENDRESS G. A.
PA (YUGG/ YU G.
PA (NIJ/ NI J.
PA (FENG/ FENG P. P.
PA (YOUN/ YOUNG P. E.
PA (GREE/ GREENE J. M.
PA (FERR/ FERRIE A. M.
PA (DUAN/ DUAN D. R.
PA (HUJ/ HU J.
PA (FLOR/ FLORENCE K. A.
PA (OLSE/ OLSEN H. S.
PA (FISC/ FISCHER C. L.
PA (EBNE/ EBNER R.
PA (BREW/ BREWER L. A.
PA (MOOR/ MOORE P. A.
PA (SHIY/ SHI Y.
PA (LAFL/ LAFLBUR D. W.
PA (LIY/ LI Y.
PA (ZENG/ ZENG Z.
PA (KYAW/ KYAW H.
XX
PI Ruben SM, Rosen CA, Soppet DR, Carter KC, Bednarik DP;
PI Endress GA, Yu G, Ni J, Feng P, Young PE, Greene JM, Ferrie AM;
PI Duan DR, Hu J, Florence KA, Olsen HS, Fischer CL, Ebner R,
PI Brewer LA, Moore PA, Shi Y, Laflaur DW, Li Y, Zeng Z, Kyaw H;
XX
DR WPI; 2003-521800/49.
DR N-PSDB; ACDB2809.
XX
XX New genes and its encoded prostate cancer antigen proteins, useful for
PT preventing, treating, ameliorating or diagnosing e.g. prostate cancers,
PT thymic hypoplasia, multiple sclerosis, AIDS, angina pectoris or cerebral
PT ischemia.
PS
PS Claim 3; SEQ ID NO 488; 260pp; English.
XX
XX The present invention relates to the isolation of novel human secreted
CC proteins and the polynucleotide sequences encoding them. The invention
CC also discloses vectors, host cells, antibodies, and recombinant methods
CC for producing human secreted proteins. The polypeptide and polynucleotide
CC sequences for the secreted proteins are useful for preventing, treating,
CC ameliorating or diagnosing medical conditions such as hyperproliferative
CC disorders (e.g. leukemia or breast cancers), wounds, reproductive
CC disorders, blood-related disorders (e.g. haemophilia or
CC thrombocytopaenia), immunodeficiencies (e.g. Wiskott-Aldrich syndrome or
CC multiple hypoplasia), autoimmune disorders (e.g. graft-versus-host disease,
CC viral or bacterial or fungal infections (e.g. AIDS or sepsis), renal
CC disorders (e.g. kidney failure), cardiovascular disorders (e.g. angina
CC pectoris, cerebral ischaemia or congenital heart defects), respiratory
CC disorders, neurological disorders (e.g. Alzheimer's disease or
CC Parkinson's disease), and inflammations (e.g. Crohn's disease). The
CC polynucleotide or polypeptide may also be used as vaccine adjuvants.
CC AB03474-AB034815 represent human secreted proteins or their fragments.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov/pspsidentry.html
XX
SQ Sequence 339 AA:
Query Match 6.9%; Score 220; DB 6; Length 339;
Best Local Similarity 22.0%; Pred. No. 6.2e-13;
Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

QY 38 GEGYRNTEYVOMGQLVPCHTNNGELSYGQLKWDITTCASERRAG-----CPCVT 91
Db 3 GPDVETPSAIQICIKMRBDDANVAGNVHGITLMIIEAGATISTRHCNSONGERCVAAL 62

QY 92 ASMDIYFEHTISVGVYNIKAKVNRAFNSMEVGIOVASEDLCSEKQWNVCAALATFVA 151
Db 63 ARVERTDFLSPMCIGEVAHSAKITYSKSKSVEQVWVMSENILTGAKKLTNRKTLWVP 122

QY 152 RREITKYKLOIIP-----RTEEF-----KMSHSAVERRRRLVYADITIKOLLANCAIQGD 203
Db 123 LSLKNVDKVLLEVPVVSXKXOEERGRKRYEAOKEEMETKWRNGDIVQPLN----- 175

QY 204 LESRDCSRWPAEKT-----RVESVELVLPPhANQGTFGGQIWMAMENVTATIASRLCR 259
Db 176 -----PEPVTYSYSSSLIHLVGPEDCTIAGFVHGCVTKMLMDVAGIYAARHCK 225

QY 260 AHPTLKAIEHFHRRGSPQVDRVLTKAIVNNAFGRSMVEGVCEAYRQEAETHRRH-INS 318
Db 226 TNIVTASVDAINFHDKIRKGCVITISGRMFTSNKSMIEIVLVADPVVDSQKRYRAAS 285

QY 319 AFMTFVVLADDDPQILPWIRPQDGERRYRREASAR 355
Db 286 AFPTYVSLSOEGRSLPVQIWPETEDKKRFEERKGR 322

RESULT 15
ADA55399
ID ADA55399 standard; protein; 370 AA.
XX
AC ADA55399;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 2967.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
XX 21-MAR-2002; 2002EP-00006586.
XX
XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isegal T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
DR WPI; 2003-395539/38.
DR N-PSDB; ADA53760.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 14; SEQ ID NO 2967; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 370 AA:
Query Match 6.9%; Score 219; DB 6; Length 370;
Best Local Similarity 22.0%; Pred. No. 9.1e-13;

Matches 74; Conservative 64; Mismatches 163; Indels 36; Gaps 6;

```
QY      38 GEGYRNPTEVQMSQVLPECHTNOGSELVSGQLMKMIDTTACLSARHAG-----CPCVT 91
      Db      35 GPDVETPSAIOICRIMRPDANVAGNVHGGTILKMIKGAGALISTHCNSONGERCVAAL 94
QY      92 ASMDIYFEHTISVGQVNNIKAKVNRAFNNSMEVGIOVASEDLCEKONVCKALATFVA 151
      Db      95 ARVERTDPLSPMCIGEVAVHSAEITYTSKHSVEYQVNMSENILTGAKKLTKAKLTMTYVP 154
QY     152 RREITKVKLKQITP-----RTEEEKMEHSVAERRRMELVY--ADTIKDLLANCAIQGD 203
      Db     155 LSLKXVDKYLEVPPVYVSGROEBOEBEGRKRYEAKLERMETKWRNGDIVQPVLN----- 207
QY     204 LESRDCSMVPAEKT---RVESVELVLPHNHNQNTFGQIMAMMENVATIASRLCR 259
      Db     208 -----PEPNTVSYSQSSLIHVLGSPDCTLHGPFVHGVTMKLMDEVAGIVAAHCK 257
QY     260 AHPITKALEMFHFRGSPQVGDRLVLKATVNNNAFKHSMVEGVCEAYROEAETHRRH-INS 318
      Db     258 TNIVTASVDALNFDHKIRKGCVTITISGMTFTSNKSMELIYVLDADPVVDSQKRYRAAS 317
QY     319 APMTFVVLADDDQQLLPWIRPQDGERRRYREASAR 355
      Db     318 AFTYVSLSQEGRSLPVPQLVPEDEKRRFEKGGR 354
```

Search completed: August 12, 2004, 15:44:04
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 23:35:26 ; Search time 883 Seconds
(without alignments)
10330.500 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 gtggaattgccttcaat.....cctgaacctatcatgagc 1857

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3228839 seqs, 245606551 residues

Total number of hits satisfying chosen parameters: 6457678

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1857	100.0	1857	16	US-10-055-624B-1
2	1839	99.0	6359	17	US-10-055-624B-1
3	1633.6	88.0	1818	16	US-10-055-624B-3
4	1230	26.2	2699	16	US-10-055-624B-5
5	471.8	25.4	813	13	US-09-925-298-181
6	471.8	25.4	813	15	US-10-102-806-181
7	403.4	21.7	2034	13	US-10-336-472-49
8	283.6	15.3	2460	17	US-10-416-314-89
9	143.2	7.7	2242	16	US-10-108-260A-1210
10	126.6	6.8	390	9	US-09-867-701-2605
11	126.2	6.8	573	10	US-09-814-353-21183
12	124.6	6.7	572	10	US-09-814-353-17043
13	124	6.7	543	10	US-09-918-995-21270
14	117.8	6.3	452	9	US-09-864-761-2374

15	115.2	6.2	594	13	US-10-027-632-191598	Sequence 191598,
16	115.2	6.2	594	13	US-10-027-632-191599	Sequence 191599,
17	115.2	6.2	594	13	US-10-027-632-191600	Sequence 191600,
18	115.2	6.2	594	16	US-10-027-632-191598	Sequence 191598,
19	115.2	6.2	594	16	US-10-027-632-191599	Sequence 191599,
20	115.2	6.2	594	16	US-10-027-632-191600	Sequence 191600,
21	114.2	6.1	629	10	US-09-814-353-4355	Sequence 4355, Ap
22	114.2	6.1	629	10	US-09-814-353-10659	Sequence 10659, A
23	112	6.0	500	13	US-10-027-632-231207	Sequence 231207,
24	112	6.0	500	13	US-10-027-632-231208	Sequence 231208,
25	112	6.0	500	13	US-10-027-632-231209	Sequence 231209,
26	112	6.0	500	16	US-10-027-632-231207	Sequence 231207,
27	112	6.0	500	16	US-10-027-632-231208	Sequence 231208,
28	112	6.0	500	16	US-10-027-632-231209	Sequence 231209,
29	101	5.4	492	9	US-09-864-761-1743	Sequence 1743, Ap
30	100.6	5.4	1204	17	US-10-115-635-2272	Sequence 272, App
31	100.6	5.4	2252	16	US-10-104-047-140	Sequence 140, App
32	100.6	5.4	2255	16	US-10-104-047-140	Sequence 881, App
33	100.6	5.4	137870	13	US-10-351-951-1	Sequence 1, Appl
34	98.8	5.3	282	10	US-09-814-353-21609	Sequence 21609, A
35	98.8	5.3	385	10	US-09-814-353-18091	Sequence 18091, A
36	97.6	5.3	263	10	US-09-814-353-5420	Sequence 5420, Ap
37	97.6	5.3	263	10	US-09-814-353-11707	Sequence 11707, A
38	95.4	5.1	47243	13	US-10-087-192-1264	Sequence 1264, App
39	94.6	5.1	2950	16	US-10-108-260A-165	Sequence 165, App
40	89.8	4.8	428	9	US-09-864-761-3429	Sequence 3429, Ap
41	89	4.8	844	13	US-10-027-632-149777	Sequence 149777,
42	89	4.8	844	16	US-10-027-632-149777	Sequence 149777,
43	88.2	4.7	905	16	US-10-120-988-311	Sequence 311, App
44	79	4.3	621	15	US-10-156-761-5013	Sequence 5013, Ap
45	79	4.3	9025608	15	US-10-156-761-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-055-624B-1
; Sequence 1, Appl
; Application No. US2003022038A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chul, Clarissa
; APPLICANT: Goddard, Audrey D
; APPLICANT: Grimaldi, J. Christopher
; TITLE OF INVENTION: BFR COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055, 624B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/263,362
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-055-624B-1

Query Match 100.0%; Score 1857; DB 16; Length 1857;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1857; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GTGGAATTGCGCTTGAATGATCCGAATGTCGGAATCAGTCGGAGCGGCTTGCC	60
DB	1	GTGGAATTGCGCTTGAATGATCCGAATGTCGGAATCAGTCGGAGCGGCTTGCC	60
QY	61	TCTGTGTTCCACCGCATCCCGGAAGTGCCTTACGTGCGGGAACGACAGTGCC	120
DB	61	TCTGTGTTCCACCGCATCCCGGAAGTGCCTTACGTGCGGGAACGACAGTGCC	120
QY	121	ATGGCAGACGGCGGAGTCCCGGAATCCCGGAGGTGAGATGAGCCAGCTGTGCTG	180
DB	121	ATGGCAGACGGCGGAGTCCCGGAATCCCGGAGGTGAGATGAGCCAGCTGTGCTG	180

Db 121 ATGGCAGACGGGAGGATACCGGAACCCCAAGGAGTGTGAGATGAGCAGGTGCTG 180
QY 181 CCTGCGCACACCAACCAAGTGTGAGCTTGAGCGTGGGCGAGCTCTCAAGTGGATTGAC 240
Db 181 CCTGCGCACACCAACCAAGTGTGAGCTTGAGCGTGGGCGAGCTCTCAAGTGGATTGAC 240
QY 241 ACCAGCGCTTGTGCTGCGGAGAGGAGCAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 241 ACCAGCGCTTGTGCTGCGGAGAGGAGCAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 GATGACATCTATTTTGTGACACACCTTATGTGTGACCAAGTGTGATATCAAGGCCAAG 360
Db 301 GATGACATCTATTTTGTGACACACCTTATGTGTGACCAAGTGTGATATCAAGGCCAAG 360
QY 361 GTGAACCGCGCTTCACTCCAGCATGAGGTGGGCACTGAGTGGCTCCGAGGAGCTG 420
Db 361 GTGAACCGCGCTTCACTCCAGCATGAGGTGGGCACTGAGTGGCTCCGAGGAGCTG 420
QY 421 TGCTCTGAGAGAGAGTGTGACAGGCTTGAGCGCTTGAGCGCTTGAGCGCTTGAGCG 480
Db 421 TGCTCTGAGAGAGAGTGTGACAGGCTTGAGCGCTTGAGCGCTTGAGCGCTTGAGCG 480
QY 481 ATCAACCAAGTGTGAGAGTGTGACAGCGCTTGAGCGCTTGAGCGCTTGAGCGCTG 540
Db 481 ATCAACCAAGTGTGAGAGTGTGACAGCGCTTGAGCGCTTGAGCGCTTGAGCGCTG 540
QY 541 AGTGTGGGCGCTGAGCGCGCGCGCTTGAGCGCTTGAGCGCTTGAGCGCTTGAGCG 600
Db 541 AGTGTGGGCGCTGAGCGCGCGCGCTTGAGCGCTTGAGCGCTTGAGCGCTTGAGCG 600
QY 601 CTGGCGCACTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 CTGGCGCACTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 GCTGAG 720
Db 661 GCTGAG 720
QY 721 GCGCAACACTTTGGGCGCGAGATCTAGCGTGTGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 GCGCAACACTTTGGGCGCGAGATCTAGCGTGTGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 AGCGGCGCTGCTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 AGCGGCGCTGCTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 CCGTCCAGAGTGTGAG 900
Db 841 CCGTCCAGAGTGTGAG 900
QY 901 AGCAGTGTGAG 960
Db 901 AGCAGTGTGAG 960
QY 961 CACATCAACAGTGTGAG 1020
Db 961 CACATCAACAGTGTGAG 1020
QY 1021 CTGCGCTGAGATTTGAG 1080
Db 1021 CTGCGCTGAGATTTGAG 1080
QY 1081 AGAAG 1140
Db 1081 AGAAG 1140
QY 1141 CTCTCCGCTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 CTCTCCGCTGCGGCTTCAAGGCGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 TTGAAGATGCTTGTGAG 1260
Db 1201 TTGAAGATGCTTGTGAG 1260

QY 1261 CTGTACACTGTGAG 1320
Db 1261 CTGTACACTGTGAG 1320
QY 1321 GATGAG 1380
Db 1321 GATGAG 1380
QY 1381 CACTACCGGAG 1440
Db 1381 CACTACCGGAG 1440
QY 1441 ACCAGCGCTGCGGAG 1500
Db 1441 ACCAGCGCTGCGGAG 1500
QY 1501 CGGAG 1560
Db 1501 CGGAG 1560
QY 1561 AOCACCGAG 1620
Db 1561 AOCACCGAG 1620
QY 1621 TGCGCGAG 1680
Db 1621 TGCGCGAG 1680
QY 1681 GTCTCGGAG 1740
Db 1681 GTCTCGGAG 1740
QY 1741 GACGCTGTGAG 1800
Db 1741 GACGCTGTGAG 1800
QY 1801 GCCAAATCAACAG 1857
Db 1801 GCCAAATCAACAG 1857

RESULT 2

US-10-717-597-301
; Sequence 301, Application US/10717597
; Publication No. US20040110221A1
; GENERAL INFORMATION:
; APPLICANT: Burczynski, Michael E.
; APPLICANT: Wyeth
; APPLICANT: Twine, Natalie C.
; APPLICANT: Dorne, Andrew J.
; APPLICANT: Trepicchio, William L.
; APPLICANT: Stonim, Donna K.
; APPLICANT: Stover, Jennifer A.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
; FILE REFERENCE: AM101080L
; CURRENT APPLICATION NUMBER: US/10/717,597
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: US 60/459,782
; PRIOR FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: US 60/427,982
; NUMBER OF SEQ ID NOS: 4904
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 301
; LENGTH: 6359
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-717-597-301

Query Match 99.0%; Score 1839; DB 17; Length 6359;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1839; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 ATGATCCAGATGTCGGAATCACTTCGACGCGGCTTGGCTCTGTGTTCTCCAAACCGC 78
DB 72 ATGATCCAGATGTCGGAATCACTTCGACGCGGCTTGGCTCTGTGTTCTCCAAACCGC 131
QY 79 ACATCCCGGAAGTCAGCTTAACTGTCGCGGGAACGACATGTCATGTCGACGCGGGA 138
DB 132 ACATCCCGGAAGTCAGCTTAACTGTCGCGGGAACGACATGTCATGTCGACGCGGGA 191
QY 139 TACCGGAACCCCAACGAGTGCAGATGAGCAGTGTGTCGCTGCTCCACCAACCA 198
DB 192 TACCGGAACCCCAACGAGTGCAGATGAGCAGTGTGTCGCTGCTCCACCAACCA 251
QY 199 CGTGTGAGCTGAGCGTCGCGGACCTGCTCAAGTGAATTGACACACGCGCTTGCCTCC 258
DB 252 CGTGTGAGCTGAGCGTCGCGGACCTGCTCAAGTGAATTGACACACGCGCTTGCCTCC 311
QY 259 GCGGAGAGGACGCTGCTGCTGCTGTCACAGCTTCCATGATGATCTATTTTGG 318
DB 312 GCGGAGAGGACGCTGCTGCTGCTGTCACAGCTTCCATGATGATCTATTTTGG 371
QY 319 CACACCTTATGTCGACAAAGTGTGATATCAAGGCAAGGTAACCGGCGCTTCAAC 378
DB 372 CACACCTTATGTCGACAAAGTGTGATATCAAGGCAAGGTAACCGGCGCTTCAAC 431
QY 379 TCCAGATGAGGTCGAGTCATCCAGTGGCTCGAGAGACCTGTCCTGAGAAACATGG 438
DB 432 TCCAGATGAGGTCGAGTCATCCAGTGGCTCGAGAGACCTGTCCTGAGAAACATGG 491
QY 439 AATGTGTGCAAGGCTTGGCCACTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 498
DB 492 AATGTGTGCAAGGCTTGGCCACTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 551
QY 499 AAGAGATTCAGCGCGCGGACAGAAAGAGATGAGACAGTGTGCGGCTGAGACGC 558
DB 552 AAGAGATTCAGCGCGCGGACAGAAAGAGATGAGACAGTGTGCGGCTGAGACGC 611
QY 559 CGGCGCATGCGGCTTGTCTATGACAGACCATCAAGACCTCTGCGCAACCTGCGCAT 618
DB 612 CGGCGCATGCGGCTTGTCTATGACAGACCATCAAGACCTCTGCGCAACCTGCGCAT 671
QY 619 CAGGCGCATGTCGAGACAGACACTGTACCGGATGTCGCGGCTGAGAAACCTGCTG 678
DB 672 CAGGCGCATGTCGAGACAGACACTGTACCGGATGTCGCGGCTGAGAAACCTGCTG 731
QY 679 GAGAGTGTGAGGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
DB 732 GAGAGTGTGAGGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 791
QY 739 CAGATCATGAGCTGAGTGAATGTCGACCATCAAGTGCAGGCAACGCGCTCTGCGCTG 798
DB 792 CAGATCATGAGCTGAGTGAATGTCGACCATCAAGTGCAGGCAACGCGCTCTGCGCTG 851
QY 799 CACCTTACGCTGAGAGGCTATTTCACTTCCAGAGCGCTGCTCCAGAGTGCAGC 858
DB 852 CACCTTACGCTGAGAGGCTATTTCACTTCCAGAGCGCTGCTCCAGAGTGCAGC 911
QY 859 CGTCTGTGCTCAAGGCTATGTCGACCAATGCTTCAAACTATGAGAGTGTGAGTGTG 918
DB 912 CGTCTGTGCTCAAGGCTATGTCGACCAATGCTTCAAACTATGAGAGTGTGAGTGTG 971
QY 919 TGGTGTGAGGCTTCTGCGCAGAGAGGCTGAGACCCAGCGCGCTCATCAACAGTGTGCT 978
DB 972 TGGTGTGAGGCTTCTGCGCAGAGAGGCTGAGACCCAGCGCGCTCATCAACAGTGTGCT 1031
QY 979 ATGACTTTTGTGTCCTGAGACGACATGACCAAGTGTGCTGCTGCTGCTGCTGCTG 1038
DB 1032 ATGACTTTTGTGTCCTGAGACGACATGACCAAGTGTGCTGCTGCTGCTGCTGCTG 1091
QY 1039 CAGCGCGGCTATGTCGAGCGGCTGACGAGAGGCTGTCGAGAGAGAGATCCGCTG 1098
DB 1092 CAGCGCGGCTATGTCGAGCGGCTGACGAGAGGCTGTCGAGAGAGAGATCCGCTG 1151

QY 1099 GACAGGAAGTACATCGTGTCTGTAAGACAGAGAGTGCCTCTTCCGCTCCCTGGGAC 1158
DB 1152 GACAGGAAGTACATCGTGTCTGTAAGACAGAGAGTGCCTCTTCCGCTCCCTGGGAC 1211
QY 1159 CTTAGCAACAGGTCATGTCGAGTCAATTAAGTCTCTCTTGAAGATGTCGCTG 1218
DB 1212 CTTAGCAACAGGTCATGTCGAGTCAATTAAGTCTCTCTTGAAGATGTCGCTG 1271
QY 1219 AAGACAACTGGTGTGTCCTGAGATGATCAGTCCGCTGTACACTCTGAGAGAT 1278
DB 1272 AAGACAACTGGTGTGTCCTGAGATGATCAGTCCGCTGTACACTCTGAGAGAT 1331
QY 1279 GACAACTGCTCTCTTCCATGAGATGTCGATGTCATGTCGATGTCGATGTCGAT 1338
DB 1332 GACAACTGCTCTCTTCCATGAGATGTCGATGTCATGTCGATGTCGATGTCGAT 1391
QY 1339 CTGCTGTCTGTCGACCTGCTGTCGAGGCTGAGGTCGAGGTCGAGGTCGAGGTCGAG 1398
DB 1392 CTGCTGTCTGTCGACCTGCTGTCGAGGCTGAGGTCGAGGTCGAGGTCGAGGTCGAG 1451
QY 1399 CTAGTCAGAGGTCAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 1458
DB 1452 CTAGTCAGAGGTCAGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAGGTCGAG 1511
QY 1459 GGTTCACAAAGCCCGACGACTTGTGATCTGTCGCTGAGGTCGAGGTCGAGGTCGAG 1518
DB 1512 GGTTCACAAAGCCCGACGACTTGTGATCTGTCGCTGAGGTCGAGGTCGAGGTCGAG 1571
QY 1519 GGGGACCCCTATGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1578
DB 1572 GGGGACCCCTATGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1631
QY 1579 GAGTACAGACGCGGAGAGACCTCTGCTCAGGCTTCTGCTGCGCGCGAGGAGGACG 1638
DB 1632 GAGTACAGACGCGGAGAGACCTCTGCTCAGGCTTCTGCTGCGCGCGAGGAGGACG 1691
QY 1639 CTGACCAAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1698
DB 1692 CTGACCAAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1751
QY 1699 TATTCCTGGGGGCTCTGCAATCCAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1758
DB 1752 TATTCCTGGGGGCTCTGCAATCCAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1811
QY 1759 CTAGCTGAGAGACCTGAGTACTCTTAAGCAATCCCGTGGCCAAATCAACAGCGGA 1818
DB 1812 CTAGCTGAGAGACCTGAGTACTCTTAAGCAATCCCGTGGCCAAATCAACAGCGGA 1871
QY 1819 TTTGATACCTTCAAGACACTGAAACCTTATCATGAGC 1857
DB 1872 TTTGATACCTTCAAGACACTGAAACCTTATCATGAGC 1910

RESULT 3
US-10-055-624B-3
Sequence 3, Application US/10055624B
Publication No. US20030220238A1
GENERAL INFORMATION:
APPLICANT: Adams, Sean H
APPLICANT: Chui, Clarissa
APPLICANT: Goddard, Audrey D
APPLICANT: Grimaldi, J. Christopher
TITLE OF INVENTION: BFT COMPOSITIONS AND METHODS OF USE
FILE REFERENCE: 9800081-0066
CURRENT APPLICATION NUMBER: US/10/055,624B
CURRENT FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/263,362
PRIOR FILING DATE: 2002-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 1818
TYPE: DNA

ORGANISM: Homo sapiens
US-10-055-624B-3

Query Match 88.0%; Score 1633.6; DB 16; Length 1818;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1636; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
15 TCAATGATCCGAATGTCGAAATCACTGAGACGGGGCTTGCTGTGTTCTCCAA 74
19 TCAATGATCCGAATGTCGAAATCACTGAGACGGGGCTTGCTGTGTTCTCCAA 78
75 CCGCAATCCCGGAAGTCAAGCTTACGTGCGGGAAACAGACAGTGCATGACAGCGCA 134
79 CCGCAATCCCGGAAGTCAAGCTTACGTGCGGGAAACAGACAGTGCATGACAGCGCA 138
135 GGGATACCGGAACCCCAAGAGTGCAGATGACCAAGCTGTGCTGCTCCCAACCA 194
139 GGGATACCGGAACCCCAAGAGTGCAGATGACCAAGCTGTGCTGCTCCCAACCA 198
195 CCAAGCTGTGAGCTGAGCGTGGGAGAGTGTCAAGTGAATGACCAAGCGCTTGCT 254
199 CCAAGCTGTGAGCTGAGCGTGGGAGAGTGTCAAGTGAATGACCAAGCGCTTGCT 258
255 GTCCGCGAGAGAGCAAGCTGCGCCCTGTGTCAAGCTTCCATGATGATCTATT 314
259 GTCCGCGAGAGAGCAAGCTGCGCCCTGTGTCAAGCTTCCATGATGATCTATT 318
315 TGAGACACCATTAAGTGTGGAACAAGTGTGAATCAAGGCGAAGTGAACCGGCTT 374
319 TGAGACACCATTAAGTGTGGAACAAGTGTGAATCAAGGCGAAGTGAACCGGCTT 378
375 CAATCCAGCATGAGAGTGGGATCCAGTGGCTCGAGAGACCTGTGCTGGAAGA 434
379 CAATCCAGCATGAGAGTGGGATCCAGTGGCTCGAGAGACCTGTGCTGGAAGA 438
435 GTGAATGTGCAAGGCTTGCCACCTTGCTGCGCCCGCAGAGATCAACCAAGTGA 494
439 GTGAATGTGCAAGGCTTGCCACCTTGCTGCGCCCGCAGAGATCAACCAAGTGA 498
495 GCTGAAGAGATCAAGCGCCGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGA 554
499 GCTGAAGAGATCAAGCGCCGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGA 558
555 GCGCGCGGCGATGCGCTTGTCTATGACAGACCATCAAGAGACCTCGCGCAAC 614
559 GCGCGCGGCGATGCGCTTGTCTATGACAGACCATCAAGAGACCTCGCGCAAC 618
615 CAATCAAGGCGATGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGAGAGACCG 674
619 CAATCAAGGCGATGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGAGAGACCG 678
675 TGTGAGAGTGTGAGAGTGTGCTTCCCTCCCAAGCAATCAACGAGGCAACCTTTGG 734
679 TGTGAGAGTGTGAGAGTGTGCTTCCCTCCCAAGCAATCAACGAGGCAACCTTTGG 738
735 GAGCAGATCAATGAGCTGATGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGAG 794
739 GAGCAGATCAATGAGCTGATGAGAGAGAGAGAGAGATGAGCAACATGTGGCGCTGAG 798
795 TGCCCAACCTTACGCTGAAGGCGATTAATGTTCCACTTCGAGGCGCGTCCAGTGG 854
799 TGCCCAACCTTACGCTGAAGGCGATTAATGTTCCACTTCGAGGCGCGTCCAGTGG 858
855 CGACCGTGTGTGCTCAAGAGCATGTGAACATGCTTCAACATCAATGAGAGTGG 914
859 CGACCGTGTGTGCTCAAGAGCATGTGAACATGCTTCAACATCAATGAGAGTGG 918
915 CGTGTGCTGAGAGGCTTATCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 974
919 CGTGTGCTGAGAGGCTTATCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 978
975 CTTATGACCTTTGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1034
```

```
DB 979 CTTATGACCTTTGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038
QY 1035 GCGCCAGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1094
DB 1039 GCGCCAGCGCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1098
QY 1095 CCTGAGACGGAATGATCATGTGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1154
DB 1099 CCTGAGACGGAATGATCATGTGCTGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
QY 1155 GAGCCTTACCAACGAGTGTACCTGAGCTACATTAACGCTCTCTCTGAAATGCTTGT 1214
DB 1159 GAGCCTTACCAACGAGTGTACCTGAGCTACATTAACGCTCTCTCTGAAATGCTTGT 1218
QY 1215 GCGCAAGAGCAACTGGGTGCTGCTCTCGAGATCAAGTCAAGTGTGCTGTACACTGTGA 1274
DB 1219 GCGCAAGAGCAACTGGGTGCTGCTCTCGAGATCAAGTCAAGTGTGCTGTACACTGTGA 1278
QY 1275 GAGTGAACAAGTTCCTCTCTCTCAATGAGAGATGAGTGGATGAGAGAGAGAGAGAG 1334
DB 1279 GAGTGAACAAGTTCCTCTCTCTCAATGAGAGATGAGTGGATGAGAGAGAGAGAGAG 1338
QY 1335 CTTCCTGCTGCTCTGAGACCTGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
DB 1339 CTTCCTGCTGCTCTGAGACCTGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1398
QY 1395 GAGCTAAGTGCAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
DB 1399 GAGCTAAGTGCAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
QY 1455 CGAGAGTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
DB 1459 CGAGAGTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
QY 1515 CAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
DB 1519 CAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
QY 1575 GCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
DB 1579 GCGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
QY 1635 CGAGCTACCAAGTCTGCT 1654
DB 1639 CGAGCTACCAAGTATCTT 1658
```

RESULT 4

US-10-055-624B-5
; Sequence 5, Application US/10055624B
; Publication No. US20030220238A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Sean H
; APPLICANT: Chui, Clarissa
; APPLICANT: Goddard, Audrey D
; APPLICANT: Grimaldi, J. Christopher
; TITLE OF INVENTION: SEIT COMPOSITIONS AND METHODS OF USE
; FIDELITY REFERENCE: 9800081-0066
; CURRENT APPLICATION NUMBER: US/10/055, 624B
; PRIOR FILING DATE: 2002-01-22
; PRIOR FILING DATE: 2002-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patencin version 3.2
; SEQ ID NO 5
; LENGTH: 2699
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-055-624B-5

Query Match 66.2%; Score 1230; DB 16; Length 2699;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;

QY 18 AATATATCCAGATATGCGAAATACCTGCGACGGGGCTTGCCCTCTGTGTTCTCAACCG 77
 Db 74 AATATATCAGATATGCGAAACCACTTGCGAAGGGGCTTGCCCTCTGTGTTCTCAATCG 133
 QY 78 CACATCCCGGAAGTCAGCTTACGTGG-----GGAAACGACAGTGTCCATGAGACGG 131
 Db 134 CACATCCCGGAAGTCATCTCCATCGAGGTCTGGAGACCTTCCTCAATGGAGAGGG 193
 QY 132 CGAGGATATCCGGAACCCACGAGATGAGATGAGCAGCTGGTGTGCTGCTCCACAC 191
 Db 194 TGAAGATATCCGGAACCCACGAGATGAGATGAGCAGCTGGTGTGCTGCTCCACAC 253
 QY 192 CAACCAACGTGTGAGCTGAGCGCTCGGACGCTGCTCAAGTGAATTGACACCAAGCTTG 251
 Db 254 CAACCAACGTGTGAGCTGAGCATTGACAGTTGCTCAAGTGAATTGACACCAAGCTTG 313
 QY 252 CCTGTCCGCGGAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 311
 Db 314 CTTATCAGCGGAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373
 QY 312 TTTTGAACACACCATTTAGTGTGGAACAAGTGTGATATCAAGGCCAAGGTGAACGGAGC 371
 Db 374 CTTGACCATATCATTATGCTGCGCCAAAGTGTGAATATCAAGGCCAAGGTGAACGGAGC 433
 QY 372 CTTTAACTCAGATGAGAGGTGGGATCCAGTGGCTCGAAGGACCTGTGCTGAGAA 431
 Db 434 CTTTAACTCAGATGAGAGGTGGGATCCAGTGGCTCGAAGGACCTGTGCTGAGAA 493
 QY 432 GCACTGGAATGTGTGCAAGGCTTGCGACCTTCTGTGCTGCTGCTGCTGCTGCTGCTG 491
 Db 494 GCACTGGAATGTGTGCAAGGCTTGCGACCTTGTGCTGCTGCTGCTGCTGCTGCTG 553
 QY 492 GAAGCTGAACAGATCAGCGCGGACAGAAAGAGAAAGATGAGAGCAGTGTGCGGCG 551
 Db 554 GAAGCTGAACAGATCAGCGCGGACAGAAAGAGAAAGATGAGAGCAGTGTGCGGCG 613
 QY 552 TGAAGCGCGGCGATGCGCTTGTGTATGAGACACCATCAAGGACCTGTGCGCACTG 611
 Db 614 TGAAGCGCGGCGATGCGCTTGTGTATGAGACACCATCAAGGACCTGTGCGCACTG 673
 QY 612 CGCATTTCAAGGCGATGTGAGAGCAGAGCTGAGCCGATGCTGCGGCTGAGAAAGC 671
 Db 674 TGTATTCAGAGCAGATTGGA---CAAGGACGTGAGAAATATGCTGCAACCGAGAGAC 730
 QY 672 CCGTGTGAGAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731
 Db 731 CCGATGTGAGAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
 QY 732 TGGGGGCGAGATCATGAGCTGATGAGAAATGTGGCCACCATTTGACAGCCGCGCTCTG 791
 Db 791 CGGGGGGAGATCATGAGCTTGTGATGAGAAATGTGGCCACCATTTGACAGCCGCGCTCTG 850
 QY 792 CCGTGTGAGAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 851
 Db 851 TCAAGCGCCACCTTACGCTCAAGGCGATGAGATTTCCATTTCCGAGGCGCGCTCTGAGT 910
 QY 852 CGGAGACGCTGTGCTGCTCAAGGCGATGAGAAATGTGGCCACCATTTGACAGCTGAGT 911
 Db 911 GGGGGAGCGCTGTGCTGCTCAAGGCGATGAGAAATGTGGCCACCATTTGACAGCTGAGT 970
 QY 912 GGGGCGTGTGCTGAGGCGCTTATCGCCAGAGGCTGAGACCCACCGGCGGCAATCAACAG 971
 Db 971 GGGGCGTGTGCTGAGGCGCTTATCGCCAGAGGCTGAGACCCACCGGCGGCAATCAACAG 1030
 QY 972 TGTCTTTATGACCTTTGTGTCTGAGCGAGATGAGACCAAGCTTGTGCTGCTGCTGAT 1031
 Db 1031 CGCTTCATATACCTTTGTGTGTCTGAGCAAGATGAGACCAAGCTTGTGCTGCTGCTGAT 1090
 QY 1032 TCGAGCCCGAGCGCGGATGTGAGCGGCGGTACCGAGAGCGCGATGCTCGAGAAAGAT 1091
 Db 1091 TCGTCCCGAGCGCGGATGTGAGCGGCGGTATCCGAGAGCGCGATGCTCGAGAAAGAT 1150

QY 1092 CCGCTGTGAGACAGAGATGATCTGTGCTCTTAAGACAGACAGAGTCCCTCTCGTGCC 1151
 Db 1151 CCGCTGTGAGACAGAGATGATCTGTGCTCTTAAGACAGAGAGTCCCTCTGTGCTGCTG 1210
 QY 1152 CTGGGACCTTGAAGACAGAGTGTATCCTGAGCTCAATTAAGTCTCTCTTGAAGATGCT 1211
 Db 1211 CTGGGACCTTGAAGACAGAGTGTATCCTGAGCTCAATTAAGTCTCTCTTGAAGAGCT 1270
 QY 1212 TGTGGCCAGAGACACTGGGTGCTGCTGAGAGATCAGTCAAGGTCCGCTGTACACTCT 1271
 Db 1271 CATGGCCAGAGACACTGGGTGCTGCTGAGAGATCAGAGAGTCCGCTGTATCACTCT 1330
 QY 1272 GAGAGATGACAGTTCCTCTCTTCCATGAGATGTGTGATGTGATGAGAGCCCA 1331
 Db 1331 AGAAGAGAC--TTCTCTCTTCACTTGGAGATGTGTGATGATGATGATGATGATG 1387
 QY 1332 GGCCTTCTGCTGCTCTGAGAGCTGTGCTGAGAGCCGAGATGTGAGATGAGAGAG 1391
 Db 1388 GGTCTTCACTGCTGTGAGACCTGCGAGAGACAGAGTGGGACAGATTAACGGAG 1447
 QY 1392 CGTGAAGCTAGTCAAGAGGTAGACGAGACGAGCCATCTACCAAGCTCAACAGCCCTGC 1451
 Db 1448 TGTGAGCTGTGCTGAGAGTGTGATGAGATGAGATGAGATGAGATGAGATGAGATG 1507
 QY 1452 CTTGAGATGACACAAAGCCCGAGAGCTTGTGATCTGAGCTGAGAGCGGAAGCTTG 1511
 Db 1508 CTTGAGAGGAGACCAACAGCCCGAGAGCTTGTGATCTGAGCTGAGAGCGGAAGCTTG 1567
 QY 1512 TGACATGAGGAGACCCCTATGTATGCTGAGAGTGTGATGAGATGAGATGAGATGAG 1571
 Db 1568 TGACATGAGGAGACCCCTATGTATGCTGAGAGTGTGATGAGATGAGATGAGATGAG 1627
 QY 1572 GAGCGCAGATGACAGAGCGCGAGAGACCTCTGCTGAGCTTGTGCTGCTGCTGCTG 1631
 Db 1628 GAGCGCAGATGACAGAGCGCGAGAGACCTCTGCTGAGCTTGTGCTGCTGCTGCTG 1687
 QY 1632 GAGCGCAGATGACCAAG 1647
 Db 1688 GAGCGCAGATGACCAAG 1703

RESULT 5
 US-09-925-298-181
 : Sequence 181, Application US/09925298
 : Publication No. US20020039764A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 : FILE REFERENCE: PA103
 : CURRENT APPLICATION NUMBER: US/09/925,298
 : CURRENT FILING DATE: 2001-08-10
 : PRIOR APPLICATION NUMBER: PCT/US06/05881
 : PRIOR FILING DATE: 2000-03-08
 : PRIOR APPLICATION NUMBER: 60/124,270
 : NUMBER OF SEQ ID NOS: 846
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 181
 : LENGTH: 813
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: (266)
 : OTHER INFORMATION: n equals a,t,g, or c
 : NAME/KEY: misc feature
 : LOCATION: (723)
 : OTHER INFORMATION: n equals a,t,g, or c
 : NAME/KEY: misc feature
 : LOCATION: (726)
 : OTHER INFORMATION: n equals a,t,g, or c
 : NAME/KEY: misc feature
 : LOCATION: (738)

OTHER INFORMATION: n equals a,t,g, or c
US-09-925-298-181

Query Match 25.4%; Score 471.8; DB 13; Length 813;
Best Local Similarity 96.4%; Pred. No. 3e-128;
Matches 487; Conservative 5; Mismatches 12; Indels 1; Gaps 1;

QY 1162 AGCAACCGAGGTACTGCTGATGCTACATTAACGCTCTCTGGAATGCTTGTGGCCAG 1221
DB 9 AGAGACCGAGGTACTGCTGATGCTACATTAACGCTCTCTGGAATGCTTGTGGCCAG 68
QY 1222 GACAACTGGGTGCTGCTCTGAGATGCTAGTCCGCTGTACACTGTGGAGATGAC 1281
DB 69 GACAACTGGGTGCTGCTCTGAGATGCTAGTCCGCTGTACACTGTGGAGATGAC 128
QY 1282 AAGTTCTCTCTCTCTCAATGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 1341
DB 129 AAGTTCTCTCTCTCTCAATGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 188
QY 1342 CTGCTCTGAGCTGCTGCTGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 1401
DB 189 CTGCTCTGAGCTGCTGCTGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 248
QY 1402 GTGACGAGGTAGAGCA-GGACGAGCCATCTACACGCTGACGAGCCGCTTCTGAGAG 1460
DB 249 GTGACGAGGTAGAGCA-GGACGAGCCATCTACACGCTGACGAGCCGCTTCTGAGAG 308
QY 1461 TCACCAAAAGCCCGAGGACTTCTGATCTGCTGCTGAGGCGGAAAGCTTGTGACATGG 1520
DB 309 TCACCAAAAGCCCGAGGACTTCTGATCTGCTGCTGAGGCGGAAAGCTTGTGACATGG 368
QY 1521 GGAACCCCTATGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1580
DB 369 GGAACCCCTATGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
QY 1581 GTACAGACGCGAGAGAGACCTTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1640
DB 429 GTACAGACGCGAGAGAGACCTTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 488
QY 1641 GACCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
DB 489 GACCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513

RESULT 6

US-10-102-806-181
Sequence 181, Application US/10102806
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P103P1CT
CURRENT APPLICATION NUMBER: US/10/102,806
PRIOR FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 181
LENGTH: 813
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (266)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (723)
OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: misc feature
LOCATION: (726)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (738)
OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-181

Query Match 25.4%; Score 471.8; DB 15; Length 813;
Best Local Similarity 96.4%; Pred. No. 3e-128;
Matches 487; Conservative 5; Mismatches 12; Indels 1; Gaps 1;

QY 1162 AGCAACCGAGGTACTGCTGATGCTACATTAACGCTCTCTGGAATGCTTGTGGCCAG 1221
DB 9 AGAGACCGAGGTACTGCTGATGCTACATTAACGCTCTCTGGAATGCTTGTGGCCAG 68
QY 1222 GACAACTGGGTGCTGCTCTGAGATGCTAGTCCGCTGTACACTGTGGAGATGAC 1281
DB 69 GACAACTGGGTGCTGCTCTGAGATGCTAGTCCGCTGTACACTGTGGAGATGAC 128
QY 1282 AAGTTCTCTCTCTCTCAATGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 1341
DB 129 AAGTTCTCTCTCTCTCAATGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 188
QY 1342 CTGCTCTGAGCTGCTGCTGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 1401
DB 189 CTGCTCTGAGCTGCTGCTGAGATGCTGCTGATGCTGATGCTGAGCCGCTTCTG 248
QY 1402 GTGACGAGGTAGAGCA-GGACGAGCCATCTACACGCTGACGAGCCGCTTCTGAGAG 1460
DB 249 GTGACGAGGTAGAGCA-GGACGAGCCATCTACACGCTGACGAGCCGCTTCTGAGAG 308
QY 1461 TCACCAAAAGCCCGAGGACTTCTGATCTGCTGCTGAGGCGGAAAGCTTGTGACATGG 1520
DB 309 TCACCAAAAGCCCGAGGACTTCTGATCTGCTGCTGAGGCGGAAAGCTTGTGACATGG 368
QY 1521 GGAACCCCTATGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1580
DB 369 GGAACCCCTATGCTGCTGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 428
QY 1581 GTACAGACGCGAGAGAGACCTTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1640
DB 429 GTACAGACGCGAGAGAGACCTTCTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 488
QY 1641 GACCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
DB 489 GACCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513

RESULT 7

US-10-336-472-49
Sequence 49, Application US/10336472
Publication No. US20040043929A1
GENERAL INFORMATION:
APPLICANT: Anderson, David W.
APPLICANT: Ballinger, Robert A.
APPLICANT: Baumgartner, Jason C.
APPLICANT: Burgess, Catherine E.
APPLICANT: Caeman, Stacie J.
APPLICANT: Chant, John S.
APPLICANT: Berghs, Constance
APPLICANT: Gangolli, Bsha A.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Ellerman, Karen
APPLICANT: Futrak, Katarzyna
APPLICANT: Gerlach, Valerie
APPLICANT: Gilbert, Jennifer A.
APPLICANT: Gunther, Erik
APPLICANT: Gorman, Linda
APPLICANT: Guo, Xiaojia Sasha
APPLICANT: Ji, Weizhen
APPLICANT: Li, Li
APPLICANT: Liu, Xiaohong

APPLICANT: Miller, Charles E.
APPLICANT: Padigar, Muralidhara
APPLICANT: Rastell, Inca
APPLICANT: MacDougall, John R.
APPLICANT: Mishra, Vishnu
APPLICANT: Pena, Carol E.A.
APPLICANT: Spaderna, Steven K.
APPLICANT: Shimkels, Richard A.
APPLICANT: Smytek, Kimberly A.
APPLICANT: Stone, David J.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Ort, Tatiana
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Tcherev, Velizar T.
APPLICANT: Verne, Corine A.M.
APPLICANT: Wolenc, Adam R.
APPLICANT: Zernusen, Bryan D.
APPLICANT: Zhong, Mei
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-533C
CURRENT APPLICATION NUMBER: US/10/336,472
CURRENT FILING DATE: 2003-01-03
PRIOR APPLICATION NUMBER: 09/746,491
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 10/005,041
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 10/023,681
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: 10/024,212
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: 10/055,569
PRIOR FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: 10/080,334
PRIOR FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: 10/092,900
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: 10/136,826
PRIOR FILING DATE: 2002-05-01
PRIOR APPLICATION NUMBER: 10/236,417
PRIOR FILING DATE: 2002-09-06
PRIOR APPLICATION NUMBER: 60/345,092
PRIOR FILING DATE: 2002-01-04
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 230
SOFTWARE: Curesqlist version 0.1
SEQ ID NO 49
LENGTH: 2034
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (61)..(1725)
US-10-336-472-49

Query Match 21.7%; Score 403.4; DB 13; Length 2034;

Best Local Similarity 56.0%; Pred. No. 5.3e-108;

Matches 849; Conservative 0; Mismatches 651; Indels 15; Gaps 4;

QY 115 AGTCCATGCGAGACGGGAGATACCGGACCCGACGAGGTGACATGAGCCAGCTG 174
DB 43 AGGCTCTCGCCCTGCGATGAGAGCGGCGCGCGCGAGGTGTCATGAGCCAGACC 102
QY 175 GTGCTCCCTGCGACACCAACCAACGCTGAGCTGCGGCGACGCTCTCAAGTG 234
DB 103 ATCAAGCGCGACGCACTGCGCGGCGAGCTGAGCGCGAGCTGCTCAAGTG 162
QY 235 ATTGACACCAAGCTGCTGCTGCTGCGGAGAGGACGAGCTGCTGCTGCTCAAGCT 294
DB 163 ATGACACCAAGCTGCTGCTGCGGCGAGCTGAGAAACATGCTGAGTTCTGCTTACGCC 222

QY 295 TCATGATGACATCTATTTTGGACACACCATTAAGTGTGGACAAAGTGTGAATATCAAG 354
DB 223 TCAGTGGATGACATACAGTTTGGAGAGACAGCTAGATGTGGAACAAGTTATTAACCATCAA 282
QY 355 GCCAAGGTGAACCGGCTTCACTCCAGATGAGAGTGGGATTCAGGTGGCTCGAG 414
DB 283 GCAAAAGTTACTAGAGATTTCAGCAAGATGAGATCAAGTATCAAGGTATGATGACG 342
QY 415 GACCTGTCTGTGAGAGACAGTGAATGTGTGCAAGGCTTGGCACCTTGTGGCCCGC 474
DB 343 GATATGCTCACTGACATTTGAGAGAGCTTTTATGTGTGCTTTCTCACTTTGTAGCCAAA 402
QY 475 CGA---GAGATCAACAAAGTGAAGCTGAAGCATCAAGCCGCGACAGAGAGAGAG 531
DB 403 CCAAGTGGAAAAAGAAAGATTCAATTAACCACTACACATTTCACTGAACAAAGATAT 462
QY 532 ATGAGACAGAGTGTGGGCTGAGCGCGGCGCATGCGCTTGTCTATGACACCATC 591
DB 463 GTGAAACATATCTGGCTGCTGAGAGAGAAAGTTGATTAACAACATGAAATACCTTT 522
QY 592 AAGGACCTCTGCGCACTGGCCATTCAGGCGCATTCGAGAGACAGACTGATGCGC 651
DB 523 AACATTTTATGAGAGAAAGTGAATTTGATGATCTCATTTTGTATGAAGAGAAAGA 582
QY 652 ATGCTGCGGCTGAAGAACCCGCTGTGAGAGTGTGAGCTGCTCTGCTCCACGCGC 711
DB 583 GCGGTTTCAACAGAGGACCTCGTTCAAGAGCATTAACATGCTGCTCCACCCCATGCA 642
QY 712 AATCAACAGGCAACCTTTGGGGGCAATATATGCTGATGAGATGTGGCAAC 771
DB 643 AACCATCAAGAAATATTTGTGGCAATTAATGCGTGAAGAGAGAGAGCTGCTACT 702
QY 772 ATTGACAGCCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 831
DB 703 ATTTCGACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
QY 832 TTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
DB 763 TTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 822
QY 892 TTCAACATATGATGAGAGTGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
DB 823 TTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 882
QY 946 GAGACCCAGCGGCGCCATCAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1005
DB 883 GAGGCGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942
QY 1006 GACCAAGCCAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065
DB 943 GAAATCTCATCAAGTTTCCCAAGATCAACCATTTCAAGGATGATTTCAAGCGCTAT 1002
QY 1066 CGAGAGCCAGTGTGCGAAAGATTCGCTGCAAGAGATGATGCTGCTGCTGCTGCTGCTGCT 1125
DB 1003 CGGAGAGCTATTCAGGCAAGCGAATTCGCTGAGCAAAATATGTTATTTCCCAAAA 1062
QY 1126 CAGACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185
DB 1063 GA---AGAGTTCATCTTTCATCACTGAGATATCAAGCAAGAGGATCCCTGAGTAC 1119
QY 1186 AATAAGTCTCTCTCTTGAAGATGCTGTGGCCAGAGCAACTGGTGTCTGCTCGAG 1245
DB 1120 AGCAATGTGAGGCGCTCAAAAATACTGGCAGCCAAAAGGGGTTGGAGGTTTACAGACT 1179
QY 1246 ATCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
DB 1180 GTGAAAAGATTAATAATATCTGGAAGACATGATTTATCTGTTGGTTGAA 1239
QY 1306 ATGATGTGATGTGATGAGAGCCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
DB 1240 AAGCATGTGGAAGTCAAGCACTTTGGCTTATGCTCTTGTCTGATCTTTAACAAGCA 1299
QY 1366 CAGAGTGGGACAAAGCATTCAGGAGCTGAGCTAGTGAAGAGTGAAGAGAGAGCAGC 1425

```

Db      1300 CCTTGGGAGAGCCCGCATTTTGTGTCCCTGTGAAGTCAATGCTGGTGAAGTGAATGATCAT 1359
QY      1426 GCCATCTACACAGCTGACACAGCCCTGCGCTCGAGGTCAACAAGCCCGAGACTTCTGTG 1485
Db      1360 CAGCTGTATACATCACTGCTCTACTAGTATGATGAC---AAACCAAGACTTGGTA 1416
QY      1486 ATCTGTGCTTGAGAGCGGAGACCTTGTGACAAATGGGGACCCCTATGTATGTGGCTGAGG 1545
Db      1417 GTACTCGTATACAGAAAGAAACCCCTCAAGATGGTAACTTTACACAGTGGCAGTGAAG 1476
QY      1546 TCGGTCAAGCTGCGCACACACCGAGAGAGCGCAGAGTACAGACGCGAGACCCCTGCGC 1605
Db      1477 TCGGTCAATTTTGGCATCGGTGCTCCCGCTCTCCACAGTACATCAAGAGTAATCATATGT 1536
QY      1606 TCAGGCTTGTGCTGCTC 1620
Db      1537 GCCGGAATTTCTCATC 1551

RESULT 8
US-10-416-314-89/c
; Sequence 89, Application US/10416314
; Publication No. US20040082508A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry
; APPLICANT: YAO, Monique G.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: CHAWLA, Nandinder K.
; APPLICANT: SANJANWALA, Madhusudan M.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LU, Yan
; APPLICANT: GIETZEN, Kimberly J.
; APPLICANT: BUREFORD, Neil
; APPLICANT: DING, Li
; APPLICANT: HARALIA, April J.A.
; APPLICANT: TANG, Y. Tom
; APPLICANT: BANDMAN, Olga
; APPLICANT: WARREN, Bridget A.
; APPLICANT: HONGHELL, Cynthia D.
; APPLICANT: LU Dyrung, Aina M.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: LEE, Sally
; APPLICANT: XU, Yuming
; APPLICANT: YANG, Junming
; APPLICANT: LAL, Preeti G.
; APPLICANT: TRAN, Bao
; APPLICANT: ISON, Craig H.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: KAREHT, Stephanie K.
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PI-0287 USN
; CURRENT APPLICATION NUMBER: US/10/416,314
; CURRENT FILING DATE: 2003-05-08
; PRIOR APPLICATION NUMBER: US 60/247,505
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: US 60/249,642
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,824
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,824
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/254,305
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/256,448
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PERL Program
; SEQ ID NO 89
; LENGTH: 2460

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 71947526CB1
US-10-416-314-89

Query Match      15.3%; Score 283.6; DB 17; Length 2460;
Best Local Similarity 98.6%; Pred. No. 1e-72;
Matches 286; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1365 GCCAGATGGGAGCAAGCACTACCGGAGCGTGAAGTGAAGAGAGTGAAGAGAGCA 1424
Db      2460 GCCAGATGGGAGCAAGCACTACCGGAGCGTGAAGTGAAGAGAGTGAAGAGAGCA 2401
QY      1425 CGCCATCTACACAGCTCACAGCCCTGCTGAGAGTCAACAAAGCCCGAGACTTGT 1484
Db      2400 CGCCATCTACACAGCTCACAGCCCTGCTGAGAGTCAACAAAGCCCGAGACTTGT 2341
QY      1485 GATCCTGCGCTCGAGGCGGAGAGCCCTGTGACAAATGGGGACCCCTATGTCATCGGCTGAG 1544
Db      2340 GATCCTGCGCTCGAGGCGGAGAGCCCTGTGACAAATGGGGACCCCTATGTCATCGGCTGAG 2281
QY      1545 GTGCGTACAGCTGCGCCACACACCGAGAGAGCGCCAGAGTACAGACGCGGAGAGACCCCTGTG 1604
Db      2280 GTGCGTACAGCTGCGCCACACACCGAGAGAGCGCCAGAGTACAGAGCGCGGAGAGACCCCTGTG 2221
QY      1605 CTCAGGCTTCTGCTCTGTGCGCGCGAGGCGGAGACCGAGTACCAAGTCTGCT 1654
Db      2220 CTCAGGCTTCTGCTCTGTGCGCGCGAGGCGGAGACCGAGTACCAAGTATCTT 2171

RESULT 9
US-10-108-260A-1210
; Sequence 1210, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1210
; LENGTH: 2242
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1210

Query Match      7.7%; Score 143.2; DB 16; Length 2242;
Best Local Similarity 85.5%; Pred. No. 2.2e-31;
Matches 183; Conservative 0; Mismatches 28; Indels 3; Gaps 2;

QY      1646 AGTGTGTGGGTAGAGGTCTCCCTGACTGAGCTGTCTCGCAAGTGAAGTCTTATTCCT 1705
Db      466 AGTGTGTGGGTAGAGGTCTCCCTGACTGAGCTGTCTCGCAAGTGAAGTCTTATTCCT 524
QY      1706 GGGGGCTCGAATCCAGGTCAAGGGTCCGAGAGCGACGTTG--GAATGAAAACCTAGC 1763
Db      525 GGGGGCTCGAATCCAGGTCAAGGGTCCGAGAGGTGTTGGAACATGGAACATGGAACCTAGC 584
QY      1764 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGGTGGCCAAATCAACAGCGGATTTGG 1823
Db      585 TGGAGGACACCTGAGTACTCTTAAAGCAATCCCGGTGGCCAAATCAATACCAATTTGG 644
QY      1824 ATACCTCAAGACACCTGAAAACCTTATCATGAGC 1857
Db      645 ATACCTCAAGACACCTGAAAACCTTATCATGAGC 678

RESULT 10
US-09-867-701-2605/c

```

```
; Sequence 2605, Application US/09867701
; Patent No. US2002013237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2605
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-701-2605
```

```
Query Match 6.8%; Score 126.6; DB 9; Length 390;
Best Local Similarity 87.9%; Pred. No. 9.9e-27;
Matches 138; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
QY 1644 CAGTGTCTGGTTAGGTTCTCCCTGACTGAGCTGTCTCGCAAGTGGCTTCTATTC 1703
DB 357 CTAGCTTCATTTGGTTAGAGTCTCCCGAGGAGCTGTCTCGCAAGTGGCTTCATTC 298
QY 1704 CTGGGGGCTCGAATCCAGTCTCAAGGGTTCGAGAGCGAGTGGATGGAAATAGC 1763
DB 297 GTGAGGCTCGAATCCAGTCTCAAGGGTTCGAGAGCGAGTGGATGGAAATAGC 238
QY 1764 TGAAGACACCTGAGTACTCTTAAAGCAATCCCGTG 1800
DB 237 TGAAGACACCTGAGTACTCTTAAAGCAATCCCGTG 201
```

```
RESULT 11
US-09-814-353-21183
; Sequence 21183, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21183
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559,
; LOCATION: 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571,
; LOCATION: 572, 573
```

```
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21183
```

```
Query Match 6.8%; Score 126.2; DB 10; Length 573;
Best Local Similarity 91.2%; Pred. No. 1.5e-26;
Matches 134; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
QY 1654 TGGTTAGGTTCTCCCTGAGTGTGTTCTCGCAAGTGGCTTCTATTCCTGGGGGCTC 1713
DB 202 TGGTTAGGTTCTCCCTGAGTGTGTTCTCGCAAGTGGCTTCTATTCCTGGGGGCTC 261
QY 1714 GAATCCAGTCAAGGGTTCGAGAGCGAGCTTGAATGGAAACTAGCTGAGAGAC 1773
DB 262 GAATCCAGTCAAGGGTTCGAGAGCGAGCTTGAATGGAAACTAGCTGAGAGAC 321
QY 1774 CTGAGTACTCTTAAAGCAATCCCGTG 1800
DB 322 CCAAGTACTCTTAAAGCAATCCCGTG 348
```

```
RESULT 12
US-09-814-353-17043
```

```
; Sequence 17043, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
```

```
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
```

```
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17043
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-17043
```

```
Query Match 6.7%; Score 124.6; DB 10; Length 572;
Best Local Similarity 90.5%; Pred. No. 4.3e-26;
Matches 133; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 1654 TGGTTAGGTTCTCCCTGAGTGTGTTCTCGCAAGTGGCTTCTATTCCTGGGGGCTC 1713
DB 202 TGGTTAGGTTCTCCCTGAGTGTGTTCTCGCAAGTGGCTTCTATTCCTGGGGGCTC 261
QY 1714 GAATCCAGTCAAGGGTTCGAGAGCGAGCTTGAATGGAAACTAGCTGAGAGAC 1773
DB 262 GAATCCAGTCAAGGGTTCGAGAGCGAGCTTGAATGGAAACTAGCTGAGAGAC 321
QY 1774 CTGAGTACTCTTAAAGCAATCCCGTG 1800
DB 322 CCAAGCACTCTTAAAGCAATCCCGTG 348
```

```
RESULT 13
US-09-918-995-21270
```

```
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: RastcSeq for Windows Version 4.0
; SEQ ID NO 191598
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-191598
```

```
Query Match 6.2%; Score 115.2; DB 13; Length 594;
Best Local Similarity 83.1%; Pred. No. 2.6e-23;
Matches 143; Conservative 0; Mismatches 28; Indels 1; Gaps 1;
```

```
QY 1649 GCTGCTGGGTTAGGTTCCCTGACTGAGTGTCTTG-GCAAGTGGCTTCTATTCTCTGG 1707
    |||||||
Db 2 GCCACTGGGTTAGGTTCCCTGACCGAGCTGGCTTCGCCAAGTGGCGTCCATTCGTGG 61

QY 1708 GGGCTGCAATCCAGTCAAGGCTCGCAGGAGCGACCGGTTGGAATGGAATACTAGCTGA 1767
    |||||||
Db 62 GGGCTCAATCCAGTCAAGGCTCGTGAAGCGATCTTGACACAGAAATACTAGCTGA 121

QY 1768 GGAACCTCTGATCTTTAAAGCAATCCCGTGCGCCAAAATCAACAGCCGAT 1819
    |||||||
Db 122 GGACACCGGAGTACTTTAAAGCAGTCCCATGTGTAGTAAGAAAGGGAGYT 173
```

Search completed: August 21, 2004, 06:02:57
Job time : 895 secs

Blank Sheet

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 20:19:58 ; Search time 5014 Seconds

(without alignments)
11059.846 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857

Sequence: 1 gfggaattgcccttcaaat.....cctgaaaccttatcatgagc 1857

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estm:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_estl:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vri:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1526.6	82.2	1785	29	AY415998 Homo sapi
2	1230	66.2	2740	11	AK052332 Mus muscu
3	1230	66.2	2993	11	AK048670 Mus muscu
4	1230	66.2	3635	11	AK031588 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	1211.8	65.3	2933	11	AK038546	Mus muscu
6	1129.6	60.8	1781	29	AY416000	Mus muscu
7	1081	58.2	1780	29	AY415999	Pan trogl
8	858.8	46.2	914	13	EX451348	EX451348
9	802.4	43.2	847	13	BQ223248	AGENCOURT
10	770.4	41.5	1008	13	EX332767	EX332767
11	698	37.6	700	14	CA389831	CA389831
12	636.4	34.3	751	9	AU141870	AU141870
13	604	32.5	604	14	CD722495	CD722495
14	596.6	32.1	892	13	BUI94842	AGENCOURT
15	592.6	31.9	911	12	BG774124	BG774124
16	556	29.9	971	10	BE732224	60156806
17	479	25.8	640	14	CF532390	UI-M-GHO-
18	461.8	24.9	635	13	BY749381	BY749381
19	460.4	24.8	462	13	EX492100	EX492100
20	451.8	24.3	649	10	BB621160	BB621160
21	437.6	23.6	702	14	CK300575	UI-E-BJ1-
22	434.8	23.4	2012	11	AK033663	AK033663
23	434.8	23.4	3050	11	AK004905	Mus muscu
24	434.8	23.4	3056	11	AK034622	Mus muscu
25	432.6	23.3	590	14	CB266287	1005192 H
26	429.8	23.1	700	14	CB518691	UI-M-GHO-
27	426.8	23.0	1026	10	BF780247	602103346
28	420.8	22.7	544	14	CB457406	714752 MA
29	419.8	22.6	487	10	BF603821	269418 MA
30	418.4	22.5	1208	14	CD505029	CD471-D09
31	391.6	21.1	884	13	EX843189	EX843189
32	388	20.9	556	12	EG710633	EG710633
33	381	20.5	654	10	BB650432	BB650432
34	378.8	20.4	578	10	BB632348	BB632348
35	362.2	19.5	961	10	BF134551	601784958
36	354.4	19.1	604	10	BB659297	BB659297
37	349.8	18.8	688	12	BQ001717	BQ001717
38	344	18.5	937	10	BF786458	602112992
39	341.8	18.4	635	14	CD218989	CD218989
40	320.4	17.3	744	13	EX843754	EX843754
41	316.8	17.1	350	12	BG956483	PM4-CT080
42	310	16.7	578	14	CD735167	CD735167
43	301	16.2	575	14	CB152012	CB152012
44	295.4	15.9	829	13	EX846471	EX846471
45	295	15.9	827	13	EX844834	EX844834

ALIGNMENTS

RESULT 1	AY415998	1785 bp	DNA	linear	GSS 17-DEC-2003
LOCUS	AY415998	Homo sapiens THEA gene, VIRTUAL TRANSCRIPT, partial sequence,			
DEFINITION	AY415998	Genomic survey sequence.			
ACCESSION	AY415998	GI:39771958			
VERSION	AY415998.1	GI:39771958			
KEYWORDS	GSS.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1785)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreria, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferred nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1785)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferreria, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.

COMMENT

FEATURES

source
1..1785
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1785
/gene="THEA"
/locus_tag="HCM5744"

gene

ORIGIN

Query Match 82.2%; Score 1526.6; DB 29; Length 1785;
Best Local Similarity 93.5%; Pred. No. 3.8e-299;
Matches 1529; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 19 ATGATCCAGAAATGTCGGAATCACCTGCGACGGGGCTTGGCCCTCTGTGTTCTTCCAAACCGC 78
DB |||||
1 ATGATCCAGAAATGTCGGAATCACCTGCGACGGGGCTTGGCCCTCTGTGTTCTTCCAAACCGC 60
QY 79 ACATCCCGAAGTCAGCCTTACGTGCGGGAACGACAGTGCATGCCAGACGCGGAGGA 138
DB |||||
61 ACATCCCGAAGTCAGCCTTACGTGCGGGAACGACAGTGCATGCCAGACGCGGAGGA 120
QY 139 TACCGAACCCACGAGGTGAGATGAGCGAGCTGGTCTGCCCTGCCACACCAACAA 198
DB |||||
121 TACCGAACCCACGAGGTGAGATGAGCGAGCTGGTCTGCCCTGCCACACCAACAA 180
QY 199 CGTGTGAGCTGAGGCTGGGCGAGCTGCTCAAGTGGATTGACACACAGGCTTGCCTGTCC 258
DB |||||
181 CGTGTGAGCTGAGGCTGGGCGAGCTGCTCAAGTGGATTGACACACAGGCTTGCCTGTCC 240
QY 259 GCGGAGAGCAGCTGGCTGCCCTGTGTGTCACAGCTTCCATGATGACATCTATTGTGAG 318
DB |||||
241 GCGGAGAGCAGCTGGCTGCCCTGTGTGTCACAGCTTCCATGATGACATCTATTGTGAG 300
QY 319 CACACATTAGTGTGGCAAGTGTGAATATCAAGGCCAAGTGAACCGGCGCTTCAAC 378
DB |||||
301 CACACATTAGTGTGGCAAGTGTGAATATCAAGGCCAAGTGAACCGGCGCTTCAAC 360
QY 379 TCCAGCATGAGGTGGCATCCAGTGGCTCGAGGACCTGTGCTCTGAGAAGCAGTGG 438
DB |||||
361 TCCAGCATGAGGNN 420
QY 439 AATGTGTGAAGCCCTTGCCACCTTGTGGCCCGCGAGAGATCACCAAGTGAAGCTG 498
DB |||||
421 NNN 480
QY 499 AAGCAGATCACGCCCGGACAGAGAGAGATGGAGCAGTGTGGCGCTGAGCGC 558
DB |||||
481 AAGCAGATCACGCCCGGACAGAGAGAGATGGAGCAGTGTGGCGCTGAGCGC 540
QY 559 CGGCGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTTGCCCAACTGCGCCATT 618
DB |||||
541 CGGCGCATGCGCTTGTCTATGACAGACCATCAAGGACCTCTTGCCCAACTGCGCCATT 600
QY 619 CAGGCGCATGAGAGAGAGACATGTAGCCGCTGTGGCGCTGAGAGACCCGTGTG 678
DB |||||
601 CAGGCGCATGAGAGAGAGACATGTAGCCGCTGTGGCGCTGAGAGACCCGTGTG 660
QY 679 GAGAGTGTGGAGCTGGTCTGCTCCACAGGCAATCACAGGCAACACCTTTCGGGCG 738
DB |||||
661 GAGAGTGTGGAGCTGGTCTGCTCCACAGGCAATCACAGGCAACACCTTTCGGGCG 720
QY 739 CAGATCATGCTGGATGAGAAATGTGCCACCATTTGACGACGCGGCTGTGCCGTGCC 798
DB |||||
721 CAGATCATGCTGGATGAGAAATGTGCCACCATTTGACGACGCGGCTGTGCCGTGCC 780
QY 799 CACCCTCAGCTCAAGGCCATTGAATGTTCATCTCCAGGCGCCCTCCAGGTCCGCGAC 858
DB |||||
781 CACCCTCAGCTCAAGGCCATTGAATGTTCATCTCCAGGCGCCCTCCAGGTCCGCGAC 840

QY 859 CGTCTGCTCTCAAGCCATCGTGAACAATGCTTCAAAACATAGCATGAGGTGGCGCTG 918
DB |||||
841 CGTCTGCTCTCAAGCCATCGTGAACAATGCTTCAAAACATAGCATGAGGTGGCGCTG 900
QY 919 TCGTGGAGGCTTATCGCCAGGAGGTGAGACCCAGCGGCCACATCAACAGTGTCTTT 978
DB |||||
901 TCGTGGAGGCTTATCGCCAGGAGGTGAGACCCAGCGGCCACATCAACAGTGTCTTT 960
QY 979 ATGACCTTTGTGCTCTTGACGACGATGACCCAGTGTGCTGCTTCCCTTCCGCTTCCGCCC 1038
DB |||||
961 ATGACCTTTGTGCTCTTGACGACGATGACCCAGTGTGCTGCTTCCCTTCCGCTTCCGCCC 1020
QY 1039 CAGCCCGGCGATGCTGAGCGCGGTACCGAGAGGCGAGTGCAGAAAGAGATCCGCTG 1098
DB |||||
1021 CAGCCCGGCGATGCTGAGCGCGGTACCGAGAGGCGAGTGCAGAAAGAGATCCGCTG 1080
QY 1099 GACAGGAAGTACATCGTGTCTTGAAGCAGACAGAGGTGCCCTCTCCCTTCCCTGGAC 1158
DB |||||
1081 GACAGGAAGTACATCGTGTCTTGAAGCAGACAGAGGTGCCCTCTCCCTTCCCTGGAC 1140
QY 1159 CTTAGCAACAGGTGTACTGAGCTACAATAACGTCTCTCTTGAAGATGCTTGTGGCC 1218
DB |||||
1141 CTTAGCAACAGGTGTACTGAGCTACAATAACGTCTCTCTTGAAGATGCTTGTGGCC 1200
QY 1219 AAGGCAACTGGGTGCTCTCGGAGATCAGTCAAGTCCGCTGTACACTCTGGAGGAT 1278
DB |||||
1201 AAGGCAACTGGGTGCTCTCGGAGATCAGTCAAGTCCGCTGTACACTCTGGAGGAT 1260
QY 1279 GAACAAGTTCCTCTCCATGAGATGCTGATGTGATGTCAGCCAGGCGCTTC 1338
DB |||||
1261 GAACAAGTTCCTCTCCATGAGATGCTGATGTGATGTCAGCCAGGCGCTTC 1320
QY 1339 CTGCTGCTCTCGGACCTGCTGAGGCGCAGAGTGGGCAAGCACTACCGAGCGGTGGAG 1398
DB |||||
1321 CTGCTGCTCTCGGACCTGCTGAGGCGCAGAGTGGGCAAGCACTACCGAGCGGTGGAG 1380
QY 1399 CTAGTGCAGCAGGTAGAGGAGCAGCGCATCTACCACTCAGCGCTCCAGCCCTCCCTCGA 1458
DB |||||
1381 CTAGTGCAGCAGGTAGAGGAGCAGCGCATCTACCACTCAGCGCTCCAGCCCTCCCTCGA 1440
QY 1459 GGTACACAAAGCCCGAGGACTTCGTGATCTCTGGCTCGAGCGGAAGCTTGTGACAAT 1518
DB |||||
1441 GGTACACAAAGCCCGAGGACTTCGTGATCTCTGGCTCGAGCGGAAGCTTGTGACAAT 1500
QY 1519 GGGGACCCCTATGTATCGCTGAGGTGCGTCACTGCTGCCACACACGAGAGAGCGCA 1578
DB |||||
1501 GGGGACCCCTATGTATCGCTGAGGTGCGTCACTGCTGCCACACACGAGAGAGCGCA 1560
QY 1579 GAGTACAGACGCGGAGAGACCTCTGCTCAGGCTTCTGCTGCGGCGAGGGGACCA 1638
DB |||||
1561 GAGTACAGACGCGGAGAGACCTCTGCTCAGGCTTCTGCTGCGGCGAGGGGACCA 1620
QY 1639 CTGACCAAGTGTGCT 1654
DB |||||
1621 CTGACCAAGTGTGCT 1636

RESULT 2

AK052332

LOCUS

DEFINITION

Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
library, clone:D33029M16 product: brown fat inducible thioesterase
2 mRNA, full insert sequence.

ACCESSION

AK052332

VERSION

AK052332.1 GI:26342570

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
92273253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, F., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2740)

ADACHI, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saico, R., Saichoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 2740
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D330029M16"
/db_xref="WGI:2421401"
/db_xref="taxon:10090"

FEATURES
source

CDs

/clone="D330029M16"
/tissue type="heart"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
132..1916
/note="unnamed protein product; brown fat inducible
thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%,
match=2641)
putative"
/codon_start=1
/protein_id="BAC34942.1"
/db_xref="GI:26342571"
/translation="MIQNVGNLRRGPFASMFNSRKSISHPSGDPTMAEGEYR
NPTEVOMSLVLPCHTNHRCGLSTIGLLKWLDTTACLSAERHAGCPCTVASMDDIYD
HTISGVGVNNAKYNAPFNSMEVGIQVVSDELCEKQMSVCALATFVAFVHRELKSV
KLKQIPLTEEEKTEHGAARMRUVIADTKILLTHCVIQDLDKDCSNMVPABK
TRYSEVLVLPHPANHQGTGGQIWMNVNATIAASRLCHNAFTPLKALTEMFHFRGP
SQVDRILVLAIVNNAFKHSMVEVCVAYRQEAETQRHINSAFMTFVLDKDDQDQ
KLPIWRPQGEGERRYREASARKKIRLDRKVLSCQAEVALSVWPSQVQLSYNN
VSSIKTLMAKDNMVLSEI SEVRLYILEEDFLSFHLEMVNVDAAOVQLLSLDRRP
ENDKYRSVELVQVDEDDALYVISPALSGNTKPDFFVLASRKFCDNGDPVIAL
RSVTLPHHETPEYRGETLCSGFLMRBGDMTKVSYNQATPGFLNYVTNVSGLS
SEFYNTFKACESFLDNRNDLAPSLQL"

ORIGIN

Query Match 66.2%; Score 1230; DB 11; Length 2740;
Best Local Similarity 85.8%; Pred. No. 8.4e-239; Indels 12; Gaps 3;
Matches 1404; Conservative 0; Mismatches 220;

Qy 18 AATGATCCAGAAATGTCGGAATCACCTGCGACGGGGTTCGGCTCTGTGTTCTCAACCG 77
Db 131 AATGATTCAGAAATGTCGGAACCACTTTCGGAAGGGCTTCGCTCTATGTTCTTAATCG 190

Qy 78 CACATCCCGGAAGTACGCTTACGTGGG-----GGGAACGACAGTGCATGGCAGAGG 131
Db 191 CACATCCCGGAAGTCAATCTCCATCGGAGTCTTGAGAGACCTCTACCATGGCAGAGG 250

Qy 132 CGAGGATACCGGAACCCGAGGTCGATGACGAGCAGCAGTGGTGGCTGCCTCCACAC 191
Db 251 TGAAGGATACCGGAACCCGAGGTCGATGACGAGCAGTGGTGGCTGCCTCCACAC 310

Qy 192 CAACCAACGTTGAGTGTGAGCGTCGCGGAGCTCTCAAGTGGATTGACACACAGCGTTG 251
Db 311 CAACCAACGTTGAGTGTGAGCGTTGAGCGTTGAGCGTTGAGCGTTGAGCGTTG 370

Qy 252 CTTGTCGCGGAGAGGACGCTGCTGCCCTCTGTTGTCACAGTTCCTATGATGATCACTA 311
Db 371 CCTATCAGCGAGAGGATGCTGCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430

Qy 312 TTTTGGACACACCATTAGTGTGACAGAGTGTGAATATCAAGCCCAAGGTGACCGGGC 371
Db 431 CTTGACACATACCATTAGTGTGACAGAGTGTGAATATCAAGCCCAAGGTGACCGGGC 490

Qy 372 CTTCAACTCCAGCATGAGGTCGGGCTCCAGGTCGGCTCCGAGGACCTGTGCTCTGAGAA 431
Db 491 CTTCAACTCCAGCATGAGGTCGGGCTCCAGGTCGGCTCCGAGGACCTGTGCTCTGAGAA 550

Qy 432 GCAGTGAATGTGTGCAAGGCTTGGCCACCTTCTGTCGCCCGCGGAGAGATCACAGGT 491
Db 551 GCAGTGAATGTGTGCAAGGCTTGGCCACCTTCTGTCGCCCGCGGAGAGATCACAGGT 610

Qy 492 GAAGCTGAAGCAGATCACGCGCGGAGAGAGAGAGATGGAGCAGATGGAGCGGCGG 551
Db 611 GAAGCTGAAGCAGATCACGCGCGGAGAGAGAGAGATGGAGCAGATGGAGCGGCGG 670

Qy 552 TGAGCGCGGCGCATGCGCTTGTCTATGACAGACCATCAAGACCTCTCTGGGCAACTG 611
Db 671 TGAGCGCGGCGCATGCGCTTGTCTATGACAGACCATCAAGACCTCTCTGGGCAACTG 730

Qy 612 CGCCATTCAGGCGGATCTGGAGAGCAGAGACTGTAGCCGCGATGTCGCGGCTGAGAGAC 671
Db 731 TGTATCCAGGACGATTGGA---CAAGGACTGCAGCAATATGTCGCCAGCGGAGAGAC 787

Qy	672	CGTGTGGAGAGTGTGAGCTGTGCTCTCCCA	CGCAATACACAGGGCAACACCTT	731
Db	788	CCGAGTGGAGAGTGTGAGCTGTGCTCTCCCTCCTCCTCAGCCAAATCATCAGGGCAATACCTT		847
Qy	732	TGGGGGCCAGATCATGCGCTTGGATGGAGAAATGTGGCCACCATTGCAGCAGCGGCTCTG	791	
Db	848	CGGGGGACAGATCATGCGCTTGGATGGAGAAATGTGGCCACCATTTGCAGCCAGCCGGCTCTG	907	
Qy	792	CCGTGCCCCACCTTACGTGTGAAGCCATTTGAAATGTTCCATCTCCGAGGCCGCTGCCAGGT	851	
Db	908	TCACGCCCACCTTACGTGTCAAGGCCATTCGAGATTTCCATTTCCGAGGCCGCTCTCAGGT	967	
Qy	852	CGGCGACCGCTGTGCTCTCAAAGCCATCTGTGAACAATGCCCTTCAAAATAGCATGTGAGGT	911	
Db	968	GGGGGACCGTCTGTGCTCAAGCCATCTGTGAATPACGCCCTTTAAGCACAGATGTGAGGT	1027	
Qy	912	GGGCGTGTGCGTGGAGCCCTATCCCGAGGAGTGTAGACCCACCGCGGCCACATCAACAG	971	
Db	1028	GGGTGTGTGTGTGGAGCGGTACCGCCAGGAAGCTGTAGACCCAGCGCCGGCACATCAACAG	1087	
Qy	972	TGCCTTTATGACCTTTCTGTGTCCTGTGAGCGAGATGACAGCCCCAGTTGTGCCCTGGAT	1031	
Db	1088	CGCTTTCATGACCTTCTGTGTCCTGTGACAAAGATGACAGCCTCAGAAGTGTGCCCTGGAT	1147	
Qy	1032	TCGGCCCCAGCCCGCGCATGTTGTAGCGCGGTATCCGAGAGCCAGTGTCCCAAGAAGAT	1091	
Db	1148	TCGTCCCCAGCCTGGAGAGGTGAACGGCGCATACCGAGAAAGCCAGTGTCCAGGAAGAAT	1207	
Qy	1092	CCGCTGTGACAGGAAGTACATCGTGTCTCTGTATGACACAGAGGTGCCCTCTCCGTCCTC	1151	
Db	1208	CCGCTGTGACAGGAATATCTTGTGTCTGTATGACAGGAAGTGGCCCTGTCTGTCTCC	1267	
Qy	1152	CTGGGACCTTAGCAACACAGGTGTACCTGAGCTACAATAAGCTCTCTCTTGAAGATGCT	1211	
Db	1268	CTGGGACCTTAGCAACACAGGTATACCTGAGCTACTACAACGTGTCTCTCTGAAGACGCT	1327	
Qy	1212	TGTGGCCACAGCAACTGGGTGTGCTCTCGAGATCAGTCAAGTCCGCTGTGACACTCT	1271	
Db	1328	CATGGCCAGGACAACTGGGTGTGCTCCGTGGAGATCAGCGAGTCCGCTGTGACATCT	1387	
Qy	1272	GGAGGTATGACAAGTTCCTCTCTTCCACATGGAGATGGTGGTCATGTGGATGCGAGCCCA	1331	
Db	1388	AGAAGAGGAC---TTCTCTCTCTTCACTTTGGAGATGCTGTAAATGTGGATGCGGCCCA	1444	
Qy	1332	GGCTTCTCTCTCTCGAGCCCTGCTCAGAGCCAGAGTGGGACAGCACTACTCGGAG	1391	
Db	1445	GGTCTTTCAGTGTCTGTACAGCTTCCGACAGAGCCAGAGTGGGACAAAGCATTTACCGGAG	1504	
Qy	1392	CGTGGAGCTATGACAGAGGTAGACAGGAGCAGCCCATCTACACATGTCCAGGCCCTGC	1451	
Db	1505	TGTGGAGCTGTGCAGCAAGTGGATGAGGATGACGCCATCTACACGTCTATCAGCCCCGC	1564	
Qy	1452	CCTCGGAGGTACACAAGCCCGAGACTTCTGTATCTCGGCTCGAGGGCGGAGCCTTG	1511	
Db	1565	CCTGAGCGGMAACCAAGCCCCAGGACTTTGTGATCCTGGCTCTTAGGCGGAGCCTTG	1624	
Qy	1512	TGACAAATGGGAGCCCTATGTCTATCGCGCTGAGGTCTGCTCAGCTGTGCCCAACACCGAGA	1571	
Db	1625	TGACAAATGGGAGCCCTATGTCTATTTGCCCTGAGGTCTGCTGCTGCCACCGACCATGA	1684	
Qy	1572	GAGCCAGAGTGTACAGCGGAGAGACCTTCTGCTCAGGCTTCTGCTCTGCGCGAGGG	1631	
Db	1685	GACACCGGAATACCACTGGGGAGACTCTCTGTTCAGGCTTCTGTCTGTGGCGTGAGGG	1744	
Qy	1632	GGACCACTGACCAAG	1647	
Db	1745	GGACCACTGACTAAG	1760	

enriched library, clone:C230001C11 product:brown fat inducible
thioesterase 2 mRNA, full insert sequence.

AK048670

AK048670.1 GI:26339489

HTC; CAP trapper.

Mus musculus (house mouse)

SOURCE

ORGANISM

REFERENCE

AUTHORS

Carninci,P. and Hayashizaki,Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

AUTHORS

2

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

TITLE

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

20499374

MEDLINE

11042159

PUBMED

REFERENCE

AUTHORS

3

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitzunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

TITLE

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

20530913

MEDLINE

11076861

PUBMED

REFERENCE

AUTHORS

4

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.

TITLE

Functional annotation of a full-length mouse cDNA collection

JOURNAL

Nature 409, 685-690 (2001)

MEDLINE

11076861

PUBMED

REFERENCE

AUTHORS

5

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

MEDLINE

12099363

PUBMED

REFERENCE

AUTHORS

6

(bases 1 to 2993)

TITLE

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,C., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawaji,S., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

Location/Qualifiers
1..2993
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C230001C11"
/db_xref="MGI:2415236"
/db_xref="taxon:10090"
/clone="C230001C11"
/tissue_types="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
116..1900
/notes="unnamed protein product; brown fat inducible thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%, match=2641)
putative"
/codon_start=1
/protein_id="BAC33416.1"
/db_xref="GI:26339490"
/translation="MIONVGNHLRRGFASFMSNRTSRKSIHPESGDPPTMARGEYVR NPTESQMSQVLVLCNTHNRGELSIGQLLKWIDTACLSEAHAGCPVTASMDIYED HTISGVGVNKKAKVNFASFMSMEVGIQVSEDLCEKQSVKALATFVAHRELKSV KLQVPLTEERTEHGAABRMRLVYADVTIKOLLTHCVIQDDLDKDCSNMVPAAEK TRVESVELVLPPEHNGQNTFGGQIIMAMENVATIASRLCHAHPTLKALEMHPFRGP SOVGRDLVLKAI VNNAPFKHMEVGVCEAVRQBAETORRHINSAPFTFVLDKDDQPO KLPIRPOGEGERRRSEARSKIRLDKRVLSCKOAVALSVPWDPSPNOVYLSYN VSSIKTILMAKDNWLVSEISVRLYLIEDPFLSHLEMVNVNDAAVQFOLLSDLRBP EWDKRVSELVQVDEDDAIHYI SPALSGNTPQDFVILASRRKPCDNGDPYVAL RSVTLPHRHTPEYORGETLCSGFLWREGDQMTKVSYNQATPGFLNLYTINVSGLS SEFYNTPKACESFLDNRNDLAPSLQL"

ORIGIN

Query Match 66.2%; Score 1230; DB 11; Length 2993;
Best local similarity 85.8%; Pred. No. 8.9e-239;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;
QY 18 AATGATCCAGATGTCGGAATACCTCGCAGCGGGCTTGGCCCTCTGTCTTCTCAACCG 77
DB 115 AATGATTCAGATGTGGCAACCACTTGGAGAGGGCTTCGCCTATGTCTCTATCG 174
QY 78 CACATCCCGGAAGTCAGCTTACGTGCG-----GGAAACGACAGTGCATGGCAGCGG 131
DB 175 CACATCCCGGAAGTCATCTCCATCCGAGTCTGGAGACCTCTACCATGGCAGAGG 234
QY 132 CGAGGATACCGGAACCCGACGAGGTGCAGATGAGCCAGTGGTGTGCTGCCCTGCCAC 191
DB 235 TGAAGGATACCGGAACCCGACGAGGTGCAGATGAGCCAGTGGTGTGCTGCCCTGCCAC 294
QY 192 CAACCAAGTGTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 251
DB 295 CAACCAAGTGTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 354
QY 252 CCTGTCCCGGAGAGGACGCTGTGCTGCCCTCTGTGTCACAGCTTCCATGGATGACATCA 311
DB 355 CCTATCAGCGAGAGGACGCTGTGCTGCCCTCTGTGTCACAGCTTCTATGGATGACATCA 414
QY 312 TTTTGAGCACACATTAAGTGTGACAGTGGTGAATATCAAGCCAGGTGAACCGGGC 371
DB 415 CTTGACCATACCATTAAGTGTGCGCAAGTGGTGAATATCAAGCCAGGTGAACCGGGC 474
QY 372 CTTCAACTCCAGATGGAGTGGGATCCAGTGGCTCGGAGGACCTGTGCTCTTGAA 431
DB 475 CTTCAACTCCAGATGGAGTGGGATCCAGTGGCTCTGTGAGATCTGTGCTCTGAA 534
QY 432 GCAGTGAATGTGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGAGAGATCAACAAGT 491
DB 535 GCAGTGGAGTGTGCAAGGCTTGGCCACCTTCGTGGCCCGCGAGGAGCTCTCAAGT 594
QY 492 GAAGTGAAGCAGATCAAGCCCGGACAGAGAGAGAGATGGAGCAGTGTGGCGGC 551

DB 595 GAAGCTGAAGCAGGTTCATCCATTGACCGAGGAGAGAGACTGAACATGGGTGGCGGC 654
QY 552 TGAGCCCGGGCGATCGGCTTGTCTATGAGACACCATCAAGGACCTCTCTGCCACTG 611
DB 655 TGAGCCCGGGCGATCGGCTTGTCTATGAGACACCATCAAGGACCTCTCTGCCACTG 714
QY 612 CGCATTTCAGGGCGATCTGAGAGCAGAGACTGTAGCCGCTGTCGGCTGAGAGAC 671
DB 715 TGTTCATCCAGGACGATTGGA---CAAGGACTGCAGCAATATGTTGCCAGCCAGAGAC 771
QY 672 CCGTGTGAGAGTGTGGAGTGTGCTTCTGCTCCCTCCCAAGCAATCAACAGGCAACACTT 731
DB 772 CCGAGTGGAGAGTGTGGAGTGTGCTTCTGCTCTCTCAGCCCAATCATCAGGCAATACCTT 831
QY 732 TGGGGCCAGATCATGGCTGTGATGAGATGTGGCCACCATTTGCAGCCAGCCGGCTCTG 791
DB 832 CGGGGACAGATCATGGCTTGGATGGAGATGTGGCCACCATTTGCAGCCAGCCGGCTCTG 891
QY 792 CCGTGCACCCCTACGCTGAAGGCCATTTGAAATGTTCCACTTCCGAGGCCGCTCCAGGT 851
DB 892 TCACGCCACCCCTACGCTGAAGGCCATTTGAGATGTTCCATTTCCGAGGCCGCTCTCAGGT 951
QY 852 CGCGCACCGCTCTGCTCAAAGCCCATCGTGAACAATGTCCTTCAACATAGCATGAGGT 911
DB 952 GGGGACCGCTCTGCTCAAAGGCCCATCGTGAACAATGTCCTTAAAGCACAGCATGAGGT 1011
QY 912 GGGCGTGTGCTGGAGGCTTATCCGACGAGGCTGAGACCCACCGGCGCCACATCAACAG 971
DB 1012 GGGTGTGTGTGGAGGCTTACCGCCAGGAAGTGAAGCCAGCGCGGCGACATCAACAG 1071
QY 972 TSCCTTTATACCTTTGTGCTCTGAGCAGATGACACGCCAGTTGCTGCCCTGGAT 1031
DB 1072 CGCTTCATGACCTTCGTGCTCTGACAAAGATGACACGCTCAGAGTGCCTCGCTGAT 1131
QY 1032 TCGGCCCCAGCCCGCGATGCTGAGCGCGGTACCGAGAGGCCAGTCCAGAAAGAGAT 1091
DB 1132 TCGTCCCGACCTGGAGAGGTGNAACGCGCATACCGAGAGGCCAGTCCAGGAAGAGAT 1191
QY 1092 CCGCTCGACAGGAAGTACATCGTGTCTCTGTAAGCAGACAGAGTGCCTCTCGTCCC 1151
DB 1192 CCGCTCGACAGGAATACCTTGTGCTCTGTAAGCAGGAGAGTGCCTCTGTGCCC 1251
QY 1152 CTGGACCTTAGCAACAGGTTACCTGAGTACAAATAAAGTCTCTCTCTTGAAGATGCT 1211
DB 1252 CTGGACCTTAGCAACAGGTTACCTGAGTACCTGAGTCTCTCTCTCTCTCTCTCTCT 1311
QY 1212 TGTGGCCAAAGCAAACTGGTGTCTCTCGAGATCAGTCAAGTCCGCTGTACATCT 1271
DB 1312 CATGGCCAAAGCAAACTGGTGTCTCTCGAGATCAGCGAGTCCGCTGTACATCT 1371
QY 1272 GGAGATGACAAGTTCCT 1331
DB 1372 AGAAGAGGAC---TTCCT 1428
QY 1332 GGCCTTCT 1391
DB 1429 GGTCTTTCAGTGTCT 1488
QY 1392 CGTGGAGTCTAGTGCAGCAGGTGAGACGAGGACGAGGACGAGGACGAGGACGAGGACGAG 1451
DB 1489 TGTGGAGTGTGACGACGAGTGGATGAGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1548
QY 1452 CCTCGAGGTGACAAGAGCCCGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1511
DB 1549 CTTGAGCGGGAACCAAGAGCCCGAGGACTTTGTGATCTCTGCTCTCTCTCTCTCTCTCTCT 1608
QY 1512 TGAACAATGGGACCCCTATGTCATCGCGCTGAGTTCGCTGAGTTCGCTGAGTTCGCTGAG 1571
DB 1609 TGAACAATGGGACCCCTATGTCATTTGCCCTGAGTTCGCTGAGTTCGCTGAGTTCGCTGAG 1668
QY 1572 GACGCAAGGTACAGACGCGGAGAGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1631

Db 1669 GACACGGGAATCCACAGTGGGAGAGACTCTGTTCAGGCTTCTGTCTGTGGCGTGAGGG 1728

Qy 1632 GGACCACTGACCAAG 1647
|||||

Db 1729 GGACCACTGACTAG 1744
|||||

RESULT 4
AK031588
LOCUS
DEFINITION
Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
enriched library, clone:6030456P18 product:brown fat inducible
thioesterase 2 mRNA, full insert sequence.
AK031588
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3635)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawaji, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES
source
1. .3635
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6030456P18"
/db_xref="MGI:2394836"
/db_xref="taxon:10090"
/clone="6030456P18"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
33. .1817
/note="unnamed protein product; brown fat inducible
thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%,
match=2641)
putative"
/codon_start=1
/protein_id="BAC27464.1"
/db_xref="GI:2632741"
/translation="MIQVNGNHLRRGFASMFNRTSKSIHPSGSDPTMAEGSGYR
NPTVQMSQLVPCHTNHRGFLSGLQLKWDITTAQLSAERHAGCPVTSMDLIFD
HTISGVGVNLIKAKVNAFNSMEVIGVSDSEKQSWVCALATFVAHREUSKV
KLQVPLTEEBTEHGAERMRRLVADTILKDLTHCVIQQDLKDCSNMVAEPK
TRVSEVLPEPHANHQGTGGQIMMENVATIAASRLCHAPTLKAIMFHRGP
SQGDRIVLKAIVNNAKHSMEVGVCEYAEQAEORRHSVAFMTFVLDKDDQP
KLPIRPQGEGRYREASRKIRLDRKLVSCQAEVLSVFPWDSNOVLYSYN
VSSDKLMAKNWVLSVEISVRVLYILEEDFLSFLHVMVNDAAQVFLDLRRRP
WMDKHSVELVQVDEDDALHYHVSFALSGNTXPDQFVILASRRKPCDNGDPVYGL
RSVTLPTHETPEYQRTGCSFLWRFGDQMTKVSYNQATPGFLNYVTNVSGLS
SEFYNTFACSFLLDNRNDLAPSLQLT"

Query Match 66.2%; Score 1230; DB 11; Length 3635;
Best Local Similarity 85.8%; Pred. No. 9.9e-239;
Matches 1404; Conservative 0; Mismatches 220; Indels 12; Gaps 3;
Qy 18 AATGATCCAGATGTCGGAATACCTCGGACGGGGCTTGGCCCTCTGTGTCTTCCAAACG 77
Db |||||
32 AATGATTCAGATGTCGGAACCACTTGGGAAGGGCTTGCCTCTATGTCTCTAATCG 91
Qy 78 CACATCCCGGAGTCACGCTTACGTGGG-----GGGAAACACAGTGCATGGCAGCG 131
Db |||||
92 CACATCCCGGAGTCATCTCCATCCGAGTCTCGGAGACCCCTCTACCATGGCAGAGG 151
Qy 132 CGAGGGATACCGGAACCCACGAGGTGCAGATGAGCCAGCTGTGCTGCCCTGCCACAC 191
Db |||||
152 TGAAGGATACCGGAACCCACGAGGTGCAGATGAGCCAGCTGTGCTGCCCTGCCACAC 211
Qy 192 CAACCAACGTTGGTGTAGCTGAGCGTFCGGGAGCTGCTCAAGTGGATTGACACCGGCTG 251
Db |||||
212 CAACCAACGTTGGTGTAGCTGAGCATTTGACAGTGTCTCAAGTGGATGACACCGGCTG 271
Qy 252 CCTGTCCCGGAGGAGGACGCTGGCTGCCCTGTGTGCACAGCTTCCATGATGATCATCTA 311
Db |||||
272 CCTATCAGCGGAGGAGGATGCTGGCTGTCCCTGTGCACAGCTTCTATGATGATCATCTA 331
Qy 312 TTTTGACACACCATTTAGTGTGGCAAGTGGTGAATATCAAGGCCAAGGTGAACCGGCG 371

332	Db	 CTTCGACCATACCATTAGTGTGGCCAAAGTGGTGAATATCAAGGCCAAAGTGTAACCGGGC	391
372	Qy	 CTTCAACTCCAGCATGGAGTGGGATCCAGGTGGCTCGGAGGACCTGTGCTCTCAGAA	431
392	Db	 CTTCAACTCCAGCATGGAGTGGGAATCCAGGTGGTCTTGAGGATCTGTGCTCTCAGAA	451
432	Qy	 GCAGTGGAAATGTGTGAAGCCCTTGGCCACTTCTGTGGCCGCCGAGAGATCAACCAAGGT	491
452	Db	 GCAGTGGAGTGTCTGCAAGGCCCTTGGCCACTTGTGGCCCAACCGGAGCTCTCCAAGGT	511
492	Qy	 GAAGCTGAACGAGATCAAGCCGGCGACAGAAAGGAGAAATGGAGCAAGTGTGGCGGC	551
512	Db	 GAAGCTGAACGAGCTATCCCATTTGACCGAGGAGGAGAACTGAACATGGGTGGCGGC	571
552	Qy	 TGAGCGCGCGCATGCGCTTCTATATGAGACACCATCAAGAGACTCTCTGCGCAACTG	611
572	Db	 TGAGCGCGCGGTATGGACTGGTCTATGAGAACCATCAAGATCTCTTAACCACTG	631
612	Qy	 CGCAATTCAGGGCCATCTGAGAGACGAGACTGTAGCCGATGCTGGCGCTGAGAGAC	671
632	Db	 TGTCAATCCAGACGATTTGA--CAAGGACTGCAGCAATATGGTGCCACCGAGAGAC	688
672	Qy	 CGGTGTGGAGAGTGTGGAGCTGTGCTCTGCCAGACCCCAATCAACAGGGCAACACCTT	731
689	Db	 CCGAGTGGAGAGTGTGGAGCTGTGCTCTGCTCTCAGCCCAATCATCAGGGCAATACCTT	748
732	Qy	 TGCGGCGCCAGATCATGGCCTTGGATGGAGATGTGCCACCATTCGACGACGCGGCTCTG	791
749	Db	 CGGGGAGACATCATGGCTTGGATGGAGATGTGCCACCATTCGACGACGCGGCTCTG	808
792	Qy	 CCGTGCGCCACCTACGCTGAAGGCCATTTGAAATGTTCCACTTCGAGGCGCGCTCCAGGT	851
809	Db	 TCACGCCCACTTACGCTCAAGGCCATCGAGATGTTCCATTTCCGAGGCGCGCTCAGGT	868
852	Qy	 CGCGCACCGTCTGGTGTCTAAAGCCATCGTGGAACAATGCCTTCAACATGACATGGAGGT	911
869	Db	 GGGGACCGTCTGGTGTCTCAAGGCCATCGTGAATAACGCCCTTAAAGCACAGCATGGAGGT	928
912	Qy	 GGCGGTGTGGTGGAGCCCTATCCGAGGAGGCTGAGACCCACCGCGCGCACATCAACAG	971
929	Db	 GGGTGTGTGTGGAGCGTACCGCCAGGAGCTTGAGACCGCGCGGCACATCAACAG	988
972	Qy	 TGCCCTTTATGACCTTTTGTGCTCTGGACGAGATGACAGCGCCCGAGTTGCTGCCCTGGAT	1031
989	Db	 CGCCTTCATGACCTTTGTTGCTCTGGACAAGATGACAGCCTCAGAAGCTTGCCCTGAT	1048
1032	Qy	 TCGGCCCCAGCCCGCGCATGGTGAAGCGGGGTACCGAGAGCGCAGTGCAGAAAGAGAT	1091
1049	Db	 TCGTCCCCAGCTGGAGAGGTTGAACGGCGATACCGAGAGCCAGTGCAGGAGAAAGAT	1108
1092	Qy	 CCGCCTGGACAGGAAGTACATCGTGTCTCTGTAACGACAGAGTGCCTCTCCGTCCTC	1151
1109	Db	 CCGCTTGGACAGGAATACCTTGTGTCTGTGAAGCGAGAGTGGCCCTGTCTGTCCC	1168
1152	Qy	 CTGGGACCTTAGCAACAGGTGTACCTGAGCTACAATAAGCTCTCTCTTGAAGATGCT	1211
1169	Db	 CTGGGACCTTAGCAACAGGTATACCTGAGCTACTCAAAGTGTCTCTCTGAAGACGCT	1228
1212	Qy	 TGTGGCCAAAGAACAACTGGGTGTCTCTCGAGATCAGTCAAGTGTCCGCTGTACACTCT	1271
1229	Db	 CATGGCCAAAGACAACCTGGGTGTCTCGGTGGATCAGCGAGGTCCGCTGTACATCCT	1288
1272	Qy	 GGAGGATCAAGATTCTCTCTTCCATCGAGATGGTGTGATCTGTGGATGACGCCCA	1331
1389	Db	 AGAAAGAGAC---TTCCTCTCTTCACTTGGAGTGGTGTGAATGTGGATGCGGCCCA	1345
1332	Qy	 GGCCTCTCTGTCTCTCGGACCTTGGCTCAGAGGCCAGGTGGGACAGCACTACCGGAG	1391
1346	Db	 GGTCTTTTCAGTGTGTTCAGACCTTGGCAGAGACCGAGTGGGACAGCATTACCGGAG	1405
1392	Qy	 CGTGGAGCTAGTGACGAGGTAGACGAGGACGACGCCATCTACCAAGTCAACGACCTTGC	1451

Db	1406	TGTGGAGCTGGTGCAGCAAGTGGATGAGGATGAGCCCATCTACCACTCATCAGCCCCG	1461
QY	1452	CCTCGGAGGTCACACAAAGCCCCAGGACTTCGTGATCCTGGCTCGAGGGGAGCCTTG	1511
Db	1466	CCTGAGGGGAAACACCAAGCCCCAGGACTTGTGATCTGSCCTCTAGGCGGAGCCTTG	1525
QY	1512	TGCAATATGGGGACCCCTATATGTCATTCGGCTTGAGGTCGGTCACGTCGCCCAACACCGAGA	1571
Db	1526	TGCAATATGGGGACCCCTATATGTCATTCGGCTTGAGGTCGGTCACGTCGCCCAACCATGA	1585
QY	1572	GAGCCAGAGTACAGACGGGAGAGACCTCTCTGCTCAGGCTTCTGCTCTGGCGCGAGGG	1631
Db	1586	GACACGGGAATACCAAGTGGGAGACTCTCTGTTCAAGGCTTCTGCTGCGCTGAGGG	1645
QY	1632	GGACCAGCTGACCAAG	1647
Db	1646	GGACCAGATGACTTAAG	1661
LOCUS	AK038546	2933 bp mRNA linear	HTC 19-SEP-2003
DEFINITION	Mus musculus adult male hypothalamus cDNA, RIKEN full-length enriched library, clone:A23003M23 product:brown fat inducible thioesterase 2 mRNA, full insert sequence.		
ACCESSION	AK038546	GI:26332636	
VERSION	AK038546.1		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Yumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5		
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
REFERENCE	6 (bases 1 to 2933)		
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,		

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://phantom.gsc.riken.go.jp/.

FEATURES source

```

1. .2933
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:A230033M23"
/db_xref="MGI:2402974"
/db_xref="taxon:10090"
/clone="A230033M23"
/sex="male"
/tissue type="hypothalamus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
99. "1838
/note="unnamed protein product; brown fat inducible
thioesterase 2 mRNA (GB|AF416923, evidence: BLASTN, 100%,
match=2641)
putative"
/codon_start=1
/protein_id="BAC30036.1"
/db_xref="GI:26332637"
/translation="MFSNRTSRKSIHPESGDPPTMAGEGVNRPTEVMSQLVLPCH
TNHRELSIGQLKWI DTTACLSAERHAGCPVTASMDDIYFDHTISYGVVNIKAKV
NRAFNSMEVGIVQVSEDLCEKQWCKALATFVAHRELKVKLQVILPTESEKTE
HGVAERERRMLVADITIKLLTHCVI ODDLDKDCSNMVPAAKTRVESVELVLPHAN
HOGNTFGQIWMWENVATIAASRLCHAHPITKALEMPHFRGSPQVGDRLVLAIVNN
AFKHSMEVGVCEVROAEQORHINSAFMTFVLDDQPKLPWRPQPGGERR
YRESARKRLDRKLVSCQAEVALSVPMWPSNVYLSYINVSLLTKLWAKDNWL
SVLESEVLVLI SEDFLSFHEUMVNDAAQVQLSLDRERPEWDRHRSVELVQVQ
DEDDAIYHVSIPALSNTPKQDFVILASRRPKDNGDPYIALRSVILPHTHEPEYQ
RGTELCGFCILWREGDQMTKVSYYNQATPGFLNVTTNVSLSEFYNTFACSEFL
DNRNDLAPSLQTL"
```

ORIGIN

Query Match 65.3%; Score 1211.8; DB 11; Length 2933;
 Best Local Similarity 85.6%; Pred. No. 4.3e-235;
 Matches 1387; Conservative 0; Mismatches 222; Indels 12; Gaps 3;

QY 33 CGGAATACACCTGGAGCGGCTTGGCTCTGTGTTCTCCAAACGACATCCCGAAGTC 92
 DB 68 CTGCACCTTCTAGACAGGCGCTTGGCTCTGTGTTCTTAATCGCATCCCGAAGTC 127

QY 93 AGCCTTACGTGCGGGA-----ACGACAGTGGCATGGCAGCGGAGGATACCGGAA 146
 DB 128 AATCTCCATCGGAGTCTGGAGACCTCTTACCATGGCAGAGGTGAAGATACCGGAA 187

QY 147 CCCACGAGGTGAGATGAGCCAGTGGTGTGCTGCTGCCACCAACCAACGTTGTA 206

```

188 CCCACGAGGTGAGATGAGCCAGCTGTACTGCTGCCACACCAACCGTGGGA 247
QY 207 GCTGAGCGTGGGCGAGCTCTCAAGTGGATTGACACCGGCTTGCCTTCGCGGAGAG 266
DB 248 GCTGAGCATTGGACAGTGTCTCAAGTGGATCGACACCAAGCTGCTTATCAGCGAGAG 307
QY 267 GCACGCTGGCTGCCCTGTGTACAGCTTCCATGGATGATCATCTATTTGAGCACACCAT 326
DB 308 GATGCTGGCTGTGCCCTGGCTCACAGCTCTTATGGATGATCATCTACTTCCACCATACCAT 367
QY 327 TAGTGTGGACAAGTGGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAACTCCAGCAT 386
DB 368 TAGTGTGGCCAAGTGGTGAATATCAAGGCCAAGGTGAACCGGGCTTCAACTCCAGCAT 427
QY 387 GGAGTGGGCATCCAGGTGGCTCGAGGAGCACTGTGCTCTGAGAAGCACTGGAATGTGTG 446
DB 428 GGAGGTGGGAATCCAGGTGGTCTCTGAGGATCTGTGCTCTGAGAAGCACTGAGTGTCTG 487
QY 447 CAAAGCCTTGGCCACTTTCGTGGCCGCGGAGAGATCAACAGGTGAAGCTGAAGCAGAT 506
DB 488 CAAAGCCTTGGCCACTTTCGTGGCCGCGGAGAGTCTTCAAGGTGAAGCTGAAGCAGAT 547
QY 507 CACGCCGCGGACAGAAAGAGGAGATGGAGCACAGTGTGGCGGCTGAGCGCGGCGCAT 566
DB 548 CATCCCATTTGACCGAGGAGGAGAGACTGAACATAGGGTGGCGGCTGAGCGCGGCGTAT 607
QY 567 GCGCTTGTCTATGAGACACCATCAAGGACCTCTCTGGCCAACTGCGCCAACTGCGCCAACTGAGGGCGA 626
DB 608 GCGACTGTCTATGAGACACCATCAAGGATCTCTCTAAACCACTGTGTCTCATCAGAGCA 667
QY 627 TCTGAGAGCAGAGACTGTAGCGCATGTGCGGCTGAGAGAGACCGCTGTGGAGATGT 686
DB 668 TTGTGA---CAGGACTCGAGCAATATGTGTGCGCAGCCAGAGAGAGAGAGAGTGT 724
QY 687 GAGCTGTCTCTGCTCCCGCAATCAACAGGGCAACACTTTGGGGGCGAGATCAT 746
DB 725 GAGCTGTGTCTGCTCTCTCAGCCCAATCATCAGGSCAATACCTTCGGGGGAGACATCAT 784
QY 747 GCGCTGATGAGAAATGTGGCCACCATTTGAGCGAGCGGCTCTGCGTGCCACCGCTAC 806
DB 785 GCGTTGGATGAGAAATGTGGCCACCATTTGAGCGAGCGGCTCTGTACGCGCCACCGCTAC 844
QY 807 GCTGAAGGCCATTGAATGTTCCACTTCCGAGGCGCGCTCCAGGTGCGGAGCGCTGTGTGT 866
DB 845 GCTCAAGGCCATCGAGATGTTCCATTTCCGAGGCGCGCTCTCAGTGGGGGAGACCGTGTGT 904
QY 867 GCTCAAGGCCATCGTGAAATGCTTCAACATAGCATGAGGTGGGCGTGTGGGTGGA 926
DB 905 GCTCAAGGCCATCGTGAAATGCTTCAACATAGCATGAGGTGGGCGTGTGGGTGGA 964
QY 927 GCGCTATCGCAGAGGCTGAGACCCACCGGGCCACATCAACAGTGCCTTTATGACCTT 986
DB 965 GCGCTATCGCAGAGGCTGAGACCCACCGGGCCACATCAACAGTGCCTTTATGACCTT 1024
QY 987 TGTGTCTCTGAGCAGATGACAGCCCGCTGTGCTGCTGAGTTCGCGCCCGCGCGG 1046
DB 1025 GGTGTCTCTGACAAAGATGATCCAGCTGAGAGTGCCTTGTGTGCTGCGGAGCGCTGG 1084
QY 1047 CGATGTTGAGCGGCGGTACCGAGAGGCGAGTGCACAAAGAGATCCCGCTGGAGGAA 1106
DB 1085 AGAGGGTGAACCGCGATACCGAGAGGCGAGTGCACAAAGAGATCCCGCTGGAGGAA 1144
QY 1107 GTACATCGTGTCTGTAAAGCAGAGAGTGCCTTCTCCGCTCCCTGGGAGCCCTAGCAA 1166
DB 1145 ATACTTGTGTCTGTAAAGCAGAGAGTGCCTTCTGTCTGCGGAGCGCTTAGCAA 1204
QY 1167 CCAGGTGTACCTGAGTACTAATAAGTCTCTCTCTTGAAGATGCTTGTGGCCAGGACAA 1226
DB 1205 CCAGGTATACCTGAGTACTAATAAGTCTCTCTTCTTGAAGAGCTCTATGGCCAGGACAA 1264
QY 1227 CTGGGTGTGTCTCGAGATCAGTCAAGGTCCGCTGTATCACTCTTGGAGGATGACAGTT 1286
```

Db 1265 CTGGTGTCTGTCCTGGAGATCAGCGAGGTCCGCTCTGATCATCTAGAGAGGAC---TT 1321
QY 1287 CCTCTCCTCCACATGAGATGCTGTCATGTGGATGCAGCCAGCCCTTCGTGCTGCT 1346
Db 1322 CCTCTCCTTTCACTTGAGATGGTGGTAAATGTGGATGCGGCCAGGTCTTTAGCTGCT 1381
QY 1347 CTGGAGCTCTGCTCAGAGGCGAGAGTGGGCAACAGCACTACCGGAGCGCTGGAGCTAGTGCA 1406
Db 1382 GTACAGACTGGCAGGAGACAGAGTGGGACAGCATTAACGAGTGTGGAGTGGTGCA 1441
QY 1407 GCGAGTAGACGAGGACGAGCCATCTACACGTCACCGAGCTGCTGCGAGGTGACAC 1466
Db 1442 GCAAGTGGATGAGATGACGCGCATCTACACGTCATCAGCCCGCCCTGAGCGGGAAAC 1501
QY 1467 AAGCCCGAGACTCTGTCATCTGCTGCTGCTGAGGCGGAGCCTGTGACATGGGACCC 1526
Db 1502 CAAGCCCGAGACTTTGTGATCTCTGCTGCTGAGCGGAAGCCTTTGACATGGGACCC 1561
QY 1527 CTATGTCATCGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1586
Db 1562 CTATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1621
QY 1587 ACGCGGAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1646
Db 1622 ACGTGGGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1681
QY 1647 G 1647
Db 1682 G 1682

RESULT 6
AY416000
LOCUS Mus musculus THBA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY416000
VERSION AY416000.1 GI:39771960
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 1781)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1781)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1781
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1781
/gene="THEA"
/locus_tag="HOM5744"
gene
Query Match 60.8%; Score 1129.6; DB 29; Length 1781;

Best Local Similarity 80.2%; Pred. No. 1.6e-218;
Matches 1311; Conservative 0; Mismatches 308; Indels 16; Gaps 4;
QY 19 ATGATCAGAAATGTGGAATACCTCGAGGGGCTTGGCTCTGTGTCTTCAACGCG 78
Db 1 ATGATTGAGAAATGTGGCAACCACTTGGGAAGGGGCTTGGCTCTGTGTCTTCAACGCG 60
QY 79 ACATCCCGGAAGTACAGCTTACGTGCG-----GGGAACGACAGTGCATGGCAGAGCG 132
Db 61 ACATCCCGGAAGTACATCTCCATCCGAGTCTGGAGACCTCTACATGGCAGAGGGT 120
QY 133 GAGGATACCGGAACCCACGAGGTGCAGATGAGCCAGCTGGTGTGCTTCCCTGCCACACC 192
Db 121 GAAGATACCGGAACCCACGAGGTGCAGATGAGCCAGCTGGTGTGCTTCCCTGCCACACC 180
QY 193 AACCAAGTGTGAGCTGAGGTGCGGACAGCTGCTCAAGTGGATGAGACACCGGCTGTC 252
Db 181 AACCAAGTGTGAGCTGAGCTTGGACAGTGTGCTCAAGTGGATGAGACACCGGCTGTC 240
QY 253 CTGTCCGCGGAGAGCGCTGGCTGCCCTGTGTACAGCTTCCATGGATGATCTAT 312
Db 241 CTATCAGCGGAGAGGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 313 TTTGAGCACACCATTTAGTGTGGAACAAGTGTGTAATATCAAGCCCAAGGTGAACCGGCG 372
Db 301 TTGACCATACCATTTAGTGTGCGCAAGTGTGTAATATCAAGCCCAAGGTGAACCGGCG 360
QY 373 TTCACTCCAGCATGAGGTGGGATCCAGTGGCTCGGAGGACCTGTGCTCTGAGAG 432
Db 361 TTCACTCCAGCATGAGGTGGGATCCAGTGGCTCGGAGGACCTGTGCTCTGAGAG 420
QY 433 CAGTGAATGTGCAAGGCTTGGCCACCTTGTGCGCCCGCGAGAGATCAACCAAGTG 492
Db 421 NNGTG 480
QY 493 AAGCTGAAGCAGATCAACGCGCGGACAGAGAGAGAGAGATGGAGCAGCTGTGCGCGCT 552
Db 481 AAGCTGAAGCAGATCAACGCGCGGACAGAGAGAGAGATGGAGCAGCTGTGCGCGCT 540
QY 553 GAGCGCGGCGATGCGCTTGTCTATGAGACAGAGCTGTAGCGCAGCTGTGCGCGCTGAGAAG 612
Db 541 GAGCGCGGCGATGCGCTTGTCTATGAGACAGAGCTGTAGCGCAGCTGTGCGCGCTGAGAAG 600
QY 613 GCAATTCAGGCGCATCTGAGAGCAGAGCTGTAGCGCAGCTGTGCGCGCTGAGAAG 672
Db 601 GTCATCCAGGACGATTTGGA---CAAGAGCTGAGAGATATGTTGTCAGCGCAGAGAG 657
QY 673 CGTGTGAGAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 732
Db 658 CGAGTGGAGAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
QY 733 GGGGGCAGATCATGGCTGGATGGAGATGTGGCCACCATTCAGGCGAGCGGCTGCTGCTGCT 792
Db 718 GGGGGCAGATCATGGCTGGATGGAGATGTGGCCACCATTCAGGCGAGCGGCTGCTGCTGCT 773
QY 793 CGTGTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
Db 774 CAGCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
QY 853 GGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 912
Db 834 GGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 893
QY 913 GCGTGTGCTGAGGCGCTATTCGCGAGGCTGAGACCCACCGCGGCGCAGATCAACAGT 972
Db 894 GGTGTGTGTGAGGCGCTACCGCGAGAGCTGAGACCCAGCGCGCGCAGATCAACAGC 953
QY 973 GCCTTTATGACCTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032
Db 954 GCCTTTATGACCTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1013
QY 1033 CCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1092

Db 1014 CGTCCCGAGCTGGAGAGGTGAACGGCGATACCGAGAGCCAGTGCACGAGAAGATC 1073
QY 1093 GCCTGGAGAGAGTACATCTGTCCTAGACAGACAGAGGTGCCCTCTCCGTCCC 1152
Db 1074 CCCTGGAGAGAGATACCTGTGCTTGTAGACAGGAGAGTGGCCCTGTGTCCC 1133
QY 1153 TGGGACCTTAGCAACAGAGTACCTGAGCTACAATAACGTCTCTCTCTGAAGATGCTT 1212
Db 1134 TGGGACCTTAGCAACAGAGTACCTGAGCTACAATAACGTCTCTCTCTGAAGAGCTC 1193
QY 1213 GTGGCCAAAGGACAACTGGGTGTGCTCTCGAGATCAAGTCAAGTCCGCTGTACACTGTG 1272
Db 1194 ATGGCCAAAGGACAACTGGGTGTGCTCTCGAGATCAAGTCAAGTCCGCTGTACACTGTG 1253
QY 1273 GAGGATGACAGTCTCTCTCTTCCACATGGAGATGGTGGTGCATGTGATCGAGCCAG 1332
Db 1254 GAAGAGGAC--TTCTCTCTCTTTCACCTTGGAGATGGTGGTGAATGTGGATCCGCCAG 1310
QY 1333 GCCTTCCTGCTCTCTCGGACCTGGCTCAGAGCCAGAGTGGGACAAAGCACTACCGGAGC 1392
Db 1311 GTCTTTACGTGCTGTACAGCTGCGCAGGAGACCAAGTGGGACAAAGCACTACCGGAGT 1370
QY 1393 GTGGAGCTAGTGCACAGGTAGACAGAGGACAGCCATCTACAGTCAACGAGCCCTGCC 1452
Db 1371 GTGGAGCTGGTGCACAGGTAGATGAGGATGAGGATGAGCCATCTACAGCTCATCAGCCCGCC 1430
QY 1453 CTCGAGGTGCACAAAGCCCGAGGACTTCGTGATCTCGGCTCGGAGCGGAGGCTGT 1512
Db 1431 CTGAGCGGGAACCAAGCCCGAGGACTTTGTGATCTCTGGCTCTAGGCGGAAGCTGT 1490
QY 1513 GACATGGGAGCCCTATGTCTATCGGCTGAGTGGTCAAGTGGTCAAGTGGCCCAACACAGGAG 1572
Db 1491 GACATGGGAGCCCTATGTCTATGCTGCTGAGTGGTCAAGTGGTCAAGTGGCCCAACACAGT 1550
QY 1573 AGCCAGAGTACAGAGCGGAGAGACCTCTGCTCAGGCTTCTGCTCGGCGAGGAGG 1632
Db 1551 ACACCGGAATACCAAGTGGGAGAGCTCTCTTTCAGGCTTCTGCTGTGGGCTGAGGG 1610
QY 1633 GACCAAGTGCACCAAG 1647
Db 1611 GACCAAGTGCACCAAG 1625

RESULT 7
LOCUS AY415999
DEFINITION Pan troglodytes THEA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY415999
VERSION AY415999.1 GI:39771959
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 1780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
2 (bases 1 to 1780)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
JOURNAL
TITLE This sequence was made by sequencing genomic exons and ordering

them based on alignment.
FEATURES
source Location/Qualifiers
1..1780
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1780
/gene="THEA"
/locus_tag="HCM5744"
ORIGIN
Query Match 58.2%; Score 1081; DB 29; Length 1780;
Best Local Similarity 66.4%; Pred. No. 1.2e-208;
Matches 1087; Conservative 0; Mismatches 549; Indels 0; Gaps 0;
QY 19 ATGATCCAGAAATGTCGGAATCACCTGCGACGCGGCTTGGCTCTGTGTCTTCAACCC 78
Db 1 ATGATCCAGAAATGTCGGAATCANNTGCGACGCGNNNNNNNNNNNNNNNNNNNNNNNN 60
QY 79 ACATCCCGGAAGTCAGCCTTACGTGCGGGGAACGACAGTGCATGCCAGCGGAGGA 138
Db 61 NNN 120
QY 139 TACCGGAACCCACGAGGTGCAGATGAGCCAGTGGTGTCTGCCCTGCCACCAACCAA 198
Db 121 TACCGGAANNN 180
QY 199 CGTGTGAGCTGAGCGTCGGGAGCTGCTCAAGTGATGACACACGCGCTTGCCTGTCC 258
Db 181 NNN 240
QY 259 GCGGAGAGCGAGCTGGCTGCCCTGTGTACAGCTTCCATGCATGCATCTATTTTGAG 318
Db 241 NCGGAGAGCGAGCTGGCTGCCCTGTGTACAGCTTCCATGCATGCATNNNNNNNNNN 300
QY 319 CACACCATTTAGTTTGGACAAGTGTGTAATATCAAGGCGCAAGTGAAACCGGCGCTTCAAC 378
Db 301 NNNNNNNNNNTGTTGGACAAGTGTGTAATATCAAGGCGCAAGTGAAACCGGCGCTTCAAC 360
QY 379 TCCAGATGAGGTGGGCTCCAGTGGCTCGAGGACCTGTGCTCTGAGAAGCAGTGG 438
Db 361 TCCAGATGAGGNN 420
QY 439 AATGTGTCAAGGCTTGCACCCTTCGTGGCCCGCGAGAGATCACCAAGTGAAGCTG 498
Db 421 NNN 480
QY 499 AAGCAGATCACGCGCGGACAGAGAGAGAGATGAGACAGTGTGTGCGGCTGAGCGC 558
Db 481 AAGCAGATCACGCGCGGACAGAGAGAGAGATGAGACAGTGTGTGCGGCTGAGCGC 540
QY 559 CGGCGCATGGCGCTTGTCTATGACAGACCATCAAGGACCTCTCGGCGCACTGCGGCATT 618
Db 541 CGGCGCATGGCGCTTGTCTATGCGGACCATCAAGGACCTCTCGGCGCACTGCGGCATT 600
QY 619 CAGGCGCATGCGGAGAGCAGAGCTGTAGCGCATGTGCGGCTGAGAGAGAGAGAGAGAG 678
Db 601 CANNNNATCTGGAGAGCAGAGACTGTAGCCGCTGAGGCTGAGAGAGAGAGAGAGAGAG 660
QY 679 GAGAGTGGAGCTGCTCTGCTCCCAACGCAATCAACAGGCGCAACCTTTGGGGGCG 738
Db 661 GAGAGTGGAGCTGCTCTGCTCCCAACGCAATCAACAGGCGCAACCTTTGGGGGCG 720
QY 739 CAGATCATGGCTGATGGAGAGATGGCCACCATTTGAGCGAGCGGCTCTGCGCGTGC 798
Db 721 CAGATCATGGCTGATGGAGAGATGGCCACCATTTGAGCGAGCGGCTCTGCGCGTGC 780
QY 799 CACCTACGCTCAAGGCGCATTTGAAATGTTCCACTTCCGAGGCGGCTCCAGCTCGCGCAG 858
Db 781 CACCTACGNNNAANACCCANNNAATGTTCCACTTCCGAGGCGGCTCCAGCTCGCGCAG 840
QY 859 CGTGTGTGCTCAAGCGCATCGTGAACATGCTTCAACATAGCATGAGGTGGGCGTGT 918

QY 1425 CGCCATACACAGTCACACAGCCCTGCCCTCGAGGTACACAAAGCCCGAGACTTCGT 1484
 |||||
 Db 661 CGCCATACACAGTCACACAGCCCTGCCCTCGAGGTACACAAAGCCCGAGACTTCGT 720
 |||||
 QY 1485 GATCCTGGCTCGAGCGGAAGCCCTTGACAAATGGGACCCCTATGTATCGCGTGAG 1544
 |||||
 Db 721 GATCCTGGCTCGAGCGGAGCCCTTGTCANATGGGACCCCTATGTATCGCGTGAG 780
 |||||
 QY 1545 GTCGCTACGCTGCCACACACCGAGAGAGCCGACAGGTACACAGCGGAGAGACCCCTCG 1604
 |||||
 Db 781 GTCGCTACGCTNGCCACACACCGAGAGAGCCGACAGGTACACAGCGGAGAGACCCCTCG 840
 |||||
 QY 1605 CTCAGGCTTCGCTCGCTCGCGGAGGGGACACAGCTGACCAAGT 1648
 |||||
 Db 841 CTCAGGCTTCGCTCGCTCGCGGAGGGGACACAGCTGACCAAGT 884
 |||||

RESULT 9
 BQ223248
 LOCUS
 DEFINITION AGENCOURT_7551001_NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6066060
 5', mRNA sequence.

ACCESSION BQ223248
 VERSION BQ223248.1 GI:20404648
 KEYWORDS EST.
 SOURCE

ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTP
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM13343 row: f column: 13
 High quality sequence stop: 647.
 Location/Qualifiers

FEATURES
 source

1. 847
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6066060"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 72"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 43.2%; Score 802.4; DB 13; Length 847;
 Best Local Similarity 98.3%; Pred. No. 2.9e-152;
 Matches 820; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
 QY 162 GATGAGCAGCTGGTGTGCTGCCCTGCCACCAACCAAGTGTGAGCTGAGCTCGGCA 221
 |||||
 Db 11 GATGAGCAGCTGGTGTGCTGCCCTGCCACCAACCAAGTGTGAGCTGAGCTCGGCA 70
 |||||
 QY 222 GCTGCTCAAGTGGATTGACACACAGGCTTGCTGTCGCGAGAGGACGCTGGCTGCC 281
 |||||
 Db 71 GCTGCTCAAGTGGATTGACACACAGGCTTGCTGTCGCGAGAGGACGCTGGCTGCC 130
 |||||
 QY 282 CTGTGTACACGCTCCATGGATGACATCTATTTTGGACACACCATAGTGTGACAGT 341
 |||||

Db 131 CTGTGTACAGCTTCATGATGATCATCTATTTTGAGCACACCATTAGTGTGGACAAGT 190
 |||||
 QY 342 GGTGATATACAGGCAAGGTGAAACCGGGCTTTCACTCCAGATGGAGGTGGCATCCA 401
 |||||
 Db 191 GGTGAATATCAAGGCAAGGTGAAACCGGGCTTTCACTCCAGATGGAGGTGGCATCCA 250
 |||||
 QY 402 GGTGCGCTCGGAGGACCTGTGTCTGTGAGAAGCAGTGGAAATGTGTGCAAGGCTTGGCCAC 461
 |||||
 Db 251 GGTGCGCTCGGAGGACCTGTGTCTGTGAGAAGCAGTGGAAATGTGTGCAAGGCTTGGCCAC 310
 |||||
 QY 462 TTCTGTGCGCGCCGAGAGATCACCAAGGTGAAGCTGAAGCAGATCACGCGCGGACAGA 521
 |||||
 Db 311 TTCTGTGCGCGCCGAGAGATCACCAAGGTGAAGCTGAAGCAGATCACGCGCGGACAGA 370
 |||||
 QY 522 AGAGGAGAGATGGAGCAGTGTGGCGGCTGAGCGCGGCGCATGCGGCTTGTCTATGC 581
 |||||
 Db 371 AGAGGAGAGATGGAGCAGTGTGGCGGCTGAGCGCGGCGCATGCGGCTTGTCTATGC 430
 |||||
 QY 582 AGACACCATCAAGGACCTCTCTGGCCAACTGCGCCATTAGGGCGGATCTGGAGAGACAGA 641
 |||||
 Db 431 AGACACCATCAAGGACCTCTCTGGCCAACTGCGCCATTAGGGCGGATCTGGAGAGACAGA 490
 |||||
 QY 642 CTGTAGCGCATGTGTGCGGCTGGAAGACCCGCTGTGAGAGTGTGGAGCTGGTCTGCC 701
 |||||
 Db 491 CTGTAGCGCATGTGTGCGGCTGGAAGACCCGCTGTGAGAGTGTGGAGCTGGTCTGCC 550
 |||||
 QY 702 TCCCCACGCCCAATCACACAGGCAACACTTTGGGGGCCAGATCATGGCTTGGATGGAGAA 761
 |||||
 Db 551 TCCCCACGCCCAATCACACAGGCAACACTTTGGGGGCCAGATCATGGCTTGGATGGAGAA 610
 |||||
 QY 762 TGTGCGCCACCATTTGAGCGAGCCGCTCTGCGCTGCCACCCCTACGCTGAAGGCCATTGA 821
 |||||
 Db 611 TGTGCGCCACCATTTGAGCGAGCCGCTCTGCGCTGCCACCCCTACGCTGAAGGCCATTGA 670
 |||||
 QY 822 AATGTTCCACTTCCAGGCGCGCTCCAGGTGGCGACCGTCTGTGTGCTCAAAAGCCATCGT 881
 |||||
 Db 671 AATGTTCCACTTCCAGGCGCGCTCCAGGTGGCGACCGTCTGTGTGCTCAAAAGCCATCGT 730
 |||||
 QY 882 GAACAATGCTTCAACATAGCATGAGGTGGCGGTGTGCGTGGAGGCTATCGCAGGA 941
 |||||
 Db 731 GAACAATGCTTCAACATAGCATGAGGTGGCGGTGTGCGTGGAGGCTATCGCAGGA 790
 |||||
 QY 942 GGCTGAGACCCACCGCGCG-CCACATCAACAGTGCCTTTATGACCTTTTGTGGTCC 994
 |||||
 Db 791 GGCTGAGACCCACCGCGCGCCACATCAAGTGCCTTTATGACCTTTTGTGGGCC 844
 |||||

RESULT 10
 BX332767

LOCUS
 DEFINITION BX332767 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
 cDNA clone CS0DC015YP21 5-PRIME, mRNA sequence.
 BX332767
 ACCESSION
 VERSION BX332767.1 GI:30312153
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 1008)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 352.f For
 more information about this cluster, see
<http://www.genoscope.cns.fr/>
 cgi-bin/cluster.cgi?seq=CS0DC015CH1Q1P1&cluster=352.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D0015CH11QPL1.

FEATURES
source
Location/Qualifiers
1. .1008
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D0015P21"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dR)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 41.5%; Score 770.4; DB 13; Length 1008;
Best Local Similarity 98.0%; Pred. No. 1e-145;
Matches 836; Conservative 3; Mismatches 7; Indels 7; Gaps 6;

QY 19 ATGATCCAGATGTCGGAATACCTCGACGGGGCTTGGCTCTGTGTTCTCCACCGC 78
Db |||||
QY 118 ATGATCCAGATGTCGGAATACCTCGACGGGGCTTGGCTCTGTGTTCTCCACCGC 177
Db |||||
QY 79 ACATCCCGAAGTCAGCTTACGTGCGGGGAAACACAGTGCATGCGAGCGGAGGGA 138
Db |||||
QY 178 ACATCCCGAAGTCAGCTTACGTGCGGGGAAACACAGTGCATGCGAGCGGAGGGA 237
Db |||||
QY 139 TACCGGAACCCACGAGGTGCAGATGAGCCAGCTGTGTGCTGCCCTGCCACACCAACAA 198
Db |||||
QY 238 TACCGGAACCCACGAGGTGCAGATGAGCCAGCTGTGTGCTGCCCTGCCACACCAACAA 297
Db |||||
QY 199 CGTGTGAGCTGAGCTGCGGACGCTCTCAAGTGGATTGACACAGGCTTGCCTGTC 258
Db |||||
QY 298 CGTGTGAGCTGAGCTGCGGACGCTCTCAAGTGGATTGACACAGGCTTGCCTGTC 357
Db |||||
QY 259 GCGGAGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
Db |||||
QY 358 GCGGAGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417
Db |||||
QY 319 CACACCATTTAGTTGGAACAGTGGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 378
Db |||||
QY 418 CACACCATTTAGTTGGAACAGTGGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 477
Db |||||
QY 379 TCCAGCATGAGGTGGGATCCAGTGGCTCGAGGACCTGTGCTGTGAGAACAGTGG 438
Db |||||
QY 478 TCCAGCATGAGGTGGGATCCAGTGGCTCGAGGACCTGTGCTGTGAGAACAGTGG 537
Db |||||
QY 439 AATGTGTGCAAGGCTTTGGCCACCTTCTGTGCGCCGCGAGAGATCACCAAGTGAAGCTG 498
Db |||||
QY 538 AATGTGTGCAAGGCTTTGGCCACCTTCTGTGCGCCGCGAGAGATCACCAAGTGAAGCTG 597
Db |||||
QY 499 AAGCAGATACCGCGCGGACAGAGAGAGAGATGAGACAGTGTGGCGGCTGAGCGC 558
Db |||||
QY 598 AAGCAGATACCGCGCGGACAGAGAGAGAGATGAGACAGTGTGGCGGCTTAAACGC 657
Db |||||
QY 559 CGGCGCATGCGCTTGTCTATGACACACATCAAG-GACCTCTGCGCAACTCGCGCAT 617
Db |||||
QY 658 CGGCGCATGCGCTTGTCTATGACACACATCAAGGACCTCTGCGCAACTCGCGCAT 717
Db |||||
QY 618 TCAGGCGCATGCGCGGACGAGACTGTAGCGCATGTGTGCGCGCTGAGAGACCGCTGT 677
Db |||||
QY 718 TCAGGCGCATGCGCGGACGAGACTGTAGCGCATGTGTGCGCGCTGAGAGACCGCTGT 777
Db |||||
QY 678 GGAGAGTGTGAGCTGTGCTTGTGCTTCCCGCAATCACAGGGGCAACACCTTTGGGGG 737
Db |||||
QY 778 GGAGAGTGTGAGCTGTGCTTGTGCTTCCCGCAATCACAGGGGCAACACCTTTGGGGG 837
Db |||||
QY 738 CCAGATCATGCGCTGGATGAGATGTGGCCACCATTTGACAGCCAGCGGCTTCCCGTGC 797
Db |||||
QY 838 CCAGATCATGCGCTGGATGAGATGTGGCCACCATTTGACAGCCAGCGGCTTCTG-CGTGC 896
Db |||||
QY 798 CCACCTTACGCTGAAGGCCATTGAATGTTTCCACTTCCGAGGCGCGTCCCGAGTGGCGGA 857
Db |||||

Db 897 CCACCTACGCTGAGG-CATTGAAATGTT-CATTCCGAGGCC--GTCCAGGTGGCGCA 952
QY 858 CCGTCTGGTGTCTCAA 872
Db 953 -CGTCTGGTGTCTCAA 966

RESULT 11
CA389831 700 bp mRNA linear EST 06-NOV-2002
LOCUS cs103b04.y1 Human Retinal pigment epithelium/choroid cDNA
DEFINITION (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs103b04
5', mRNA sequence.
CA389831
VERSION CA389831.1 GI:24720367
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 700)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,
Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
Expressed sequence tag analysis of human RPE/choroid for the
NEIBank Project: Over 6000 non-redundant transcripts, novel genes
and splice variants
Mol. Vis. 8 (4), 205-220 (2002)
JOURNAL 22103460
MEDLINE 12107410
PUBMED
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 103 row: b column: 04
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1. .700
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="cs103b04"
/tissue_type="RPE/choroid"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Retinal pigment epithelium/choroid cDNA
(Un-normalized, unamplified): cs"
/notes="Organ: Eye; Vector: pCMVSPORT6; Two different donor
eyes (75-80 years old) yielded approximately 600 mg of
dissected RPE/choroid tissue. This in turn yielded 340 ug
of total RNA and 7 ug of mRNA. A directionally cloned cDNA
library in the pCMVSPORT6 vector was constructed at Life
Technologies (Rockville, MD; now part of Invitrogen Corp),
essentially following the protocols of the SuperScript
Plasmid System (Invitrogen Corp).
<http://www.invitrogen.com/>. The library code
designation was cs. For this library, cDNA inserts were
cloned into the NotI/MluI sites of the vector. EST
analysis was performed on the unamplified library at the
NIH Intramural Sequencing Center (NISC)."

ORIGIN

Query Match 37.6%; Score 698; DB 14; Length 700;
Best Local Similarity 100.0%; Pred. No. 4.1e-131;
Matches 698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 CAATCACCAGGCAACACCTTTGGGGCCAGATCATGGCTGGATGAGATGTGGCCAC 770
Db |||||
QY 3 CAATCACCAGGCAACACCTTTGGGGCCAGATCATGGCTGGATGAGATGTGGCCAC 62
Db |||||


```

QY 771 CATTGACGACGCGCTCTGCGGTGCGCCACCTACGCTGAGGCGCAATTGAAATGTTCCA 830
Db 63 CATTGACGACGCGCTCTGCGGTGCGCCACCTACGCTGAGGCGCAATTGAAATGTTCCA 122
QY 831 CTTCCGAGCGCGTCCGAGGTGCGGACCGTCTGTGTCTCAAGCAATGCTGAAACAATGC 890
Db 123 CTTCCGAGCGCGTCCGAGGTGCGGACCGTCTGTGTCTCAAGCAATGCTGAAACAATGC 182
QY 891 CTTCAAACATAGCATGGAGGTGGCGTGTGCGTGAGGCGCTATCCGAGGCGTGGAGC 950
Db 183 CTTCAAACATAGCATGGAGGTGGCGTGTGCGTGAGGCGCTATCCGAGGCGTGGAGC 242
QY 951 CCACCGCGCGCACATCAACAGTGCCTTTATGACCTTTGTGTCTGCGGCGCAGATGACCA 1010
Db 243 CCACCGCGCGCACATCAACAGTGCCTTTATGACCTTTGTGTCTGCGGCGCAGATGACCA 302
QY 1011 GCGCCAGTGTGCGCTGATTTGCGGCCGACCGCGCGGATGTGAGCGCGGTACCGAGA 1070
Db 303 GCGCCAGTGTGCGCTGATTTGCGGCCGACCGCGCGGATGTGAGCGCGGTACCGAGA 362
QY 1071 GCGCCAGTGTGCGCTGATTTGCGGCCGACCGCGCGGATGTGAGCGCGGTACCGAGA 1130
Db 363 GCGCCAGTGTGCGCTGATTTGCGGCCGACCGCGCGGATGTGAGCGCGGTACCGAGA 422
QY 1131 AGAGGTGCGCCCTCTCCGTCCTGCGACCTAGCAACCAAGTGTACCTGAGCTACAATAA 1190
Db 423 AGAGGTGCGCCCTCTCCGTCCTGCGACCTAGCAACCAAGTGTACCTGAGCTACAATAA 482
QY 1191 CGTCTCTCTTGAAGATGCTTGTGGCCAGGACAACTGGGTGCTGCTCGGAGATGAG 1250
Db 483 CGTCTCTCTTGAAGATGCTTGTGGCCAGGACAACTGGGTGCTGCTCGGAGATGAG 542
QY 1251 TCAGTGTGCGCTGATCTGCGAGGATGACAAAGTCTCTCTCTCCACATGAGATGCT 1310
Db 543 TCAGTGTGCGCTGATCTGCGAGGATGACAAAGTCTCTCTCTCCACATGAGATGCT 602
QY 1311 GGTGATGTGATGATGACGACCGCCCTCTGCTGCTCTGCGACCTGCGGTACAGGCGCAGA 1370
Db 603 GGTGATGTGATGATGACGACCGCCCTCTGCTGCTCTGCGACCTGCGGTACAGGCGCAGA 662
QY 1371 GTGGACAGACACTACCGAGGCGTGGAGTGTGAGC 1408
Db 663 GTGGACAGACACTACCGAGGCGTGGAGTGTGAGC 700

```

```

RESULT 12
LOCUS AU141870 751 bp mRNA linear EST 05-AUG-2002
DEFINITION AU141870 THYRO1 Homo sapiens cDNA clone THYRO1001374 5', mRNA
sequence.
ACCESSION AU141870
VERSION AU141870.1 GI:11003391
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 751)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and

```

```

FEATURES
source
Helix Research Institute.
Location/Qualifiers
1..751
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="THYRO1001374"
/tissue type="thyroid gland"
/clone_lib="THYRO1"
/note="Vector: pME18SFL3"
ORIGIN
Query Match 34.3%; Score 636.4; DB 9; Length 751;
Best Local Similarity 97.8%; Pred. No. 1.4e-118;
Matches 654; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
QY 19 ATGATCCAGAAATGTCGGAATACCTGCGACCGGGCTTGGCCCTCTGTGTTCTTCCAAACCGC 78
Db 83 ATGATCCAGAAATGTCGGAATACCTGCGACCGGGCTTGGCCCTCTGTGTTCTTCCAAACCGC 142
QY 79 ACATCCCGGAAGTTCAGCTTACGTGCGGGGAACGACAGTGCATGGCAGCGCGGAGGA 138
Db 143 ACATCCCGGAAGTTCAGCTTACGTGCGGGGAACGACAGTGCATGGCAGCGCGGAGGA 202
QY 139 TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGGTGTCTCCCTGCCACCAACCAA 198
Db 203 TACCGGAACCCCAACGAGGTGCAGATGAGCCAGCTGGTGTCTCCCTGCCACCAACCAA 262
QY 199 CGTGTGAGCTGAGCGTGGGCGAGCTGCTCAAGTGGATGACACACGCGCTTGCCTGTCC 258
Db 263 CGTGTGAGCTGAGCGTGGGCGAGCTGCTCAAGTGGATGACACACGCGCTTGCCTGTCC 322
QY 259 GCGGAGAGGACGCTGGCTGCCCTGTGTACAGCTTCCATGCATGACATCTATTTTGAG 318
Db 323 GCGGAGAGGACGCTGGCTGCCCTGTGTACAGCTTCCATGCATGACATCTATTTTGAG 382
QY 319 CACACCATTTAGTGTGGCAAGTGTGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 378
Db 383 CACACCATTTAGTGTGGCAAGTGTGTGAATATCAAGGCCAAGGTGAACCGGCGCTTCAAC 442
QY 379 TCCAGCATGGAGGTGGGCGATCCAGTGGCTCGGAGGACCTGTGCTCTGAGAGAGTGG 438
Db 443 TCCAGCATGGAGGTGGGCGATCCAGTGGCTCGGAGGACCTGTGCTCTGAGAGAGTGG 502
QY 439 AATGTGTGCAAGGCGCTTGGCCACCTTGTGGCGCGCGAGAGATCACCAAGGTGAAGCTG 498
Db 503 AATGTGTGCAAGGCGCTTGGCCACCTTGTGGCGCGCGAGAGATCACCAAGGTGAAGCTG 562
QY 499 AAGCAGATCAAGCGCGGACAGAGAGAGATGGAGCAGAGTGTGGCGGCTGAGCGC 558
Db 563 AAGCAGATCAAGCGCGGACAGAGAGAGATGGAGCAGAGTGTGGCGGCTGAGCGC 622
QY 559 CGGCGCATGCGCTTGTCTATCGACACCACTCAAGGACCTCTGGCCAACTGGCCATT 618
Db 623 CGGCGCATGCGCTTGTCTATCGACACCACTCAAGGACCTCTGGCCAACTGGCCATT 682
QY 619 CAGGCGCATCTGG-AGAGCAGAGACTGTAGCCCGCATGTGGCGGCTGAGAGACCGCGTGT 677
Db 683 CAGGCGCATCTGAGAAAGCANAGACTGTACCGCATGTGGCGGCTTANAAAAACCGTGT 742
QY 678 GGAGAGTGT 686
Db 743 NGAAAAATGT 751
RESULT 13
LOCUS CD722495
DEFINITION oJ10d11.Y1 Human lacrimal gland, unamplified: oJ Homo sapiens cDNA
clone oJ10d11 5', mRNA sequence.
ACCESSION CD722495
VERSION CD722495.1 GI:32273343
KEYWORDS EST.

```


SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 604)
 AUTHORS Dickinson, D., Laurie, G. and Wistow, G.
 TITLE Expressed sequence tag analysis of human lacrimal gland
 JOURNAL Unpublished (2002)
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 10 row: d column: 11
 Seq primer: M13RPI reverse primer (ABI).
 FEATURES
 source
 1..604
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="oj10d11"
 /tissue_type="lacrimal gland"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human lacrimal gland, unamplified: oj"
 /notes="Organ: Eye; Vector: pCMVSPORT6; RNA was extracted from 2 human lacrimal glands. A directionally cloned cDNA library in the pCMVSPORT6 vector(life technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual
 (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor
 (5'-pGACTAGTTCTAGATCGGAGGGCGGCC(M)15-3'). EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."
 ORIGIN
 Query Match 32.5%; Score 604; DB 14; Length 604;
 Best Local Similarity 100.0%; Pred. No. 4.6e-112;
 Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 CTGTGCTCTGAGAAGCAGTGAATGTGTGCAAGGCGCTTGGCCACCTTCGTGGCGCCGCCGA 477
 DB 1 CTGTGCTCTGAGAAGCAGTGAATGTGTGCAAGGCGCTTGGCCACCTTCGTGGCGCCGCCGA 60
 QY 478 GAGATCACCAGGTGAAGCTGAAGCAGATCACGCCGCGGACAGAGAGGAGAGATGGAG 537
 DB 61 GAGATCACCAGGTGAAGCTGAAGCAGATCACGCCGCGGACAGAGAGGAGAGATGGAG 120
 QY 538 CACAGTGTGGCGGTGAGCGCGCGCGATGCGCTTGTCTATGAGACACCATCAAGGAC 597
 DB 121 CACAGTGTGGCGGTGAGCGCGCGCGATGCGCTTGTCTATGAGACACCATCAAGGAC 180
 QY 598 CTCTGTGCCCACTGCGCAATTCAGGCGCATCTGAGAGCAGAGACTGTAGCCGATCGTG 657
 DB 181 CTCTGTGCCCACTGCGCAATTCAGGCGCATCTGAGAGCAGAGACTGTAGCCGATCGTG 240
 QY 658 CCGGCTGAGAGACCCGTGTGAGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
 DB 241 CCGGCTGAGAGACCCGTGTGAGAGTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 QY 718 CAGGGCAACACTTTGGGGGCGAGATCATGCGCTGGATGGAGATGTGGCCACCATTTGCA 777
 DB 301 CAGGGCAACACTTTGGGGGCGAGATCATGCGCTGGATGGAGATGTGGCCACCATTTGCA 360
 QY 778 GCCAGCGCGCTCTGCGGTGCGCCACCTTACCGTGAAGGCCAATTTGAAATGTTCCACTTCCGA 837
 DB 361 GCCAGCGCGCTCTGCGGTGCGCCACCTTACCGTGAAGGCCAATTTGAAATGTTCCACTTCCGA 420

QY 838 GGCCCTCCAGTCCGCGACCGCTGTGGTCTCAAGAGCCATCGTCAACATGCTTCAAA 897
 DB 421 GGCCCTCCAGTCCGCGACCGCTGTGGTCTCAAGAGCCATCGTCAACATGCTTCAAA 480
 QY 898 CATAGCATGAGAGTGGCGGTGTGCGTGGAGGCTTATGCGCAGGAGGTGAGACCCACCG 957
 DB 481 CATAGCATGAGAGTGGCGGTGTGCGTGGAGGCTTATGCGCAGGAGGTGAGACCCACCG 540
 QY 958 CGCCACATCAACAGTGGCTTTATGACCTTGTGCTGCGCAGCATGACCGACCCAG 1017
 DB 541 CGCCACATCAACAGTGGCTTTATGACCTTGTGCTGCGCAGCATGACCGACCCAG 600
 QY 1018 TTGC 1021
 DB 601 TTGC 604
 RESULT 14
 BU194842
 LOCUS AGENCOURT_7970201 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6164904
 DEFINITION 5', mRNA sequence.
 ACCESSION BU194842
 VERSION BU194842.1 GI:22708826
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 892)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: AFCC/DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM13522 row: m column: 01
 High quality sequence stop: 554.
 FEATURES
 source
 1..892
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6164904"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_72"
 /notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."
 ORIGIN
 Query Match 32.1%; Score 596.6; DB 13; Length 892;
 Best Local Similarity 97.5%; Pred. No. 1.9e-110;
 Matches 627; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
 QY 39 TCACCTCGCAGCGGGCTTGGCTCTGTGTTCTCCAACCGCACATCCCGAGTCAGCCTT 98
 DB 126 TCCCATCTTTGGGGCTTGGCTCTGTGTTCTCCAACACACATCCCGAGTCAGCCTT 185
 QY 99 ACCTGCGGGAAACGACAGTCCCATGGCAGAGGGATACCGAACCCCGAGGT 158
 DB 186 ACCTGCGGGAAACGACAGTCCCATGGCAGAGGGATACCGAACCCCGAGGT 245
 QY 159 GCAGATGAGCGAGTGGTGTGCTGCTGCCACACCAACAGTGGTGGTGGTGGTGGTGGTGG 218

Db 246 GCAGATGAGCCAGCTGGTGTCTGCTCCCTGCCACACCAACCAACAGTGTGTGAGCTGAGCGTGG 305
 Qy 219 GCAGCTGTCAAGTGATTTGACACACAGCTTGCCTGTCGCGGAGAGGACGCTGGCTG 278
 Db 306 GCAGCTGTCAAGTGATTTGACACACAGCTTGCCTGTCGCGGAGAGGACGCTGGCTG 365
 Qy 279 CCCTGTGTCAAGCTTCCATGGATGACATCTATTTTGGAGCACACCATTAGTGTGGACA 338
 Db 366 CCCTGTGTCAAGCTTCCATGGATGACATCTATTTTGGAGCACACCATTAGTGTGGACA 425
 Qy 339 AGTGTGTAATCAAGGCAAGGTGAACCGGCGCTTCACTCCAGCATGGAGGTGGCAT 398
 Db 426 AGTGTGTAATCAAGGCAAGGTGAACCGGCGCTTCACTCCAGCATGGAGGTGGCAT 485
 Qy 399 CCAGTGGCTCGGAGGACCTGTGCTCTGAGAAGCAGTGGATGTGCAAGGCGCTTGGC 458
 Db 486 CCAGTGGCTCGGAGGACCTGTGCTCTGAGAAGCAGTGGATGTGCAAGGCGCTTGGC 545
 Qy 459 CACCTTGTGGCCCGCGAGAGATCACCAAGGTGAAGCTGAAGCATCACGCCCGGAC 518
 Db 546 CACCTTGTGGCCCGCGAGAGATCACCAAGGTGAAGCTGAAGCATCACGCCCGGAC 605
 Qy 519 AGAAGAGAGAGATGGAGCAGTGTGGCGGCTGAGCGCGGCGCATGGCGCTTGTCTA 578
 Db 606 AGAAGAGAGAGATGGAGCAGTGTGGCGGCTGAGCGCGGCGCATGGCGCTTGTCTA 665
 Qy 579 TGCAGACACCATCAAGGACCTCCTGGCCAACTGGC - CCATTGAGGCGGATCTGGAGAGCA 637
 Db 666 TGCAGACACCATCAAGGACCTCCTGGCCAACTGGC - CCATTGAGGCGGATCTGGAGAGCA 725
 Qy 638 GAGACTGTAGCGGCATGTGGCC - GGCTGAGAGAGCCCGTGTGG 679
 Db 726 GAGACTGTAGCGGCATGTGGCCGGGCTGAGAGAGACCCCTGTGTG 768

RESULT 15

BG774124 911 bp mRNA linear EST 15-MAY-2001
 LOCUS 602662096F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4810631 5',
 DEFINITION mRNA sequence.

ACCESSION BG774124 GI:14044423

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE

REFERENCE
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI666 row: p column: 24

High quality sequence stop: 851.

Location/Qualifiers

1. .911

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4810631"

/tissue type="choriocarcinoma"

/lab host="DH10B (phage-resistant)"

/clone lib="NIH_MGC_21"

/note="Organ: placenta; Vector: pORF7; Site: 1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

FEATURES

source

Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 31.9%; Score 592.6; DB 12; Length 911;
 Best Local Similarity 99.3%; Pred. No. 1.2e-109;
 Matches 595; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1056 GCGGCGGTACCGAGAGCGCAGTGCAGAAAGAAAGATCCGCTCGGACAGGAGTACATGCT 1115
 Db 2 GCGGCGGTACCGAGAGCGCAGTGCAGAAAGAAAGATCCGCTCGGACAGGAGTACATGCT 61
 Qy 1116 GTCTGTAAAGCAGACAGAGGTGCCCTCTCCGTCCCTGGGACCCCTAGCAACAGGTGTA 1175
 Db 62 GTCTGTAAAGCAGACAGAGGTGCCCTCTCCGTCCCTGGGACCCCTAGCAACAGGTGTA 121
 Qy 1176 CCTGAGCTACATACCTCTCTCTTGAAGATGCTTGGCCAGGACCACTGGGTGCT 1235
 Db 122 CCTGAGCTACATACCTCTCTCTTGAAGATGCTTGGCCAGGACCACTGGGTGCT 181
 Qy 1236 GTCTCTCGAGATCAGTCAAGTCCGCTCTACACTCTGGAGATGACAAAGTTCCTCTCCTT 1295
 Db 182 GTCTCTCGAGATCAGTCAAGTCCGCTCTACACTCTGGAGATGACAAAGTTCCTCTCCTT 241
 Qy 1296 CCACATGAGATGGTGGTGCATGTGGATGAGCCAGCCCTTCTGCTGCTCGGACCT 1355
 Db 242 CCACATGAGATGGTGGTGCATGTGGATGAGCCAGCCCTTCTGCTGCTCGGACCT 301
 Qy 1356 GCGTCAGAGCGCAGATGGGACAGCACTACCGAGCGTGGAGCTAGTGCAGCAGGTAGA 1415
 Db 302 GCGTCAGAGCGCAGATGGGACAGCACTACCGAGCGTGGAGCTAGTGCAGCAGGTAGA 361
 Qy 1416 CGAGGACGACGCCCTATACACAGTCCAGGCGCTCGGAGGTACACAAAGCCCCA 1475
 Db 362 CGAGGACGACGCCCTATACACAGTCCAGGCGCTCGGAGGTACACAAAGCCCCA 421
 Qy 1476 GGACTTCGTGATCCCTGGCCTCGAGCGGAAAGCCTTTGACAAATGGGGACCCCTATGTCAT 1535
 Db 422 GGACTTCGTGATCCCTGGCCTCGAGCGGAAAGCCTTTGACAAATGGGGACCCCTATGTCAT 481
 Qy 1536 CGCGCTGAGGTCGCTCAGCTGCCACACACCGAGAGACGCCAGTACAGACGCGGAGA 1595
 Db 482 CGCGCTGAGGTCGCTCAGCTGCCACACACCGAGAGACGCCAGTACAGACGCGGAGA 541
 Qy 1596 GACCTCTGTCTCAGGCTTCTGCTCTGCGCGAGGGGACCCAGCTGACCAAGTGTGCT 1654
 Db 542 GACCTCTGTCTCAGGCTTCTGCTCTGCGCGAGGGGACCCAGCTGACCAAGTGTGCT 600

Search completed: August 21, 2004, 00:59:07

Job time : 5026 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2004, 21:09:05 ; Search time 150 Seconds

(without alignments)
6870.292 Million cell updates/sec

Title: US-10-055-624B-1

Perfect score: 1857
Sequence: 1 ggggaatgcctttcaaat.....cctgaacattatcatgagc 1857

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.6	3.5	504	4	US-09-252-991A-15035
2	60.4	3.3	939	4	US-09-252-991A-1680
3	58.4	3.1	609	4	US-09-252-991A-2523
4	58.4	3.1	1518	4	US-09-252-991A-2393
5	56.6	3.0	480	4	US-09-252-991A-1630
6	56.2	3.0	1509	4	US-09-149-476-179
7	56.2	3.0	1530	4	US-09-149-476-306
8	53	2.9	1098	4	US-09-252-991A-14407
9	53	2.9	1566	4	US-09-252-991A-14264
10	53	2.9	1908	4	US-09-252-991A-14353
11	50.6	2.7	1647	3	US-09-328-111-475
12	49.2	2.6	975	4	US-09-252-991A-5935
13	49.2	2.6	2196	4	US-09-252-991A-5953
14	49.2	2.6	3036	4	US-09-252-991A-5934
15	48	2.6	270	4	US-09-252-991A-14378
16	47.8	2.6	48328	4	US-09-536-002-27
17	46.8	2.5	2588	4	US-09-480-017-3
18	46	2.5	489	4	US-09-540-236-716
19	45.6	2.5	1190	6	5223414-1
20	45.4	2.4	1432	4	US-09-620-312D-420
21	45	2.4	1131	4	US-09-252-991A-6825
22	45	2.4	2463	4	US-09-252-991A-6775
23	44.2	2.4	1308	4	US-09-252-991A-914
24	44.2	2.4	1449	4	US-09-252-991A-1050
25	44	2.4	1182	4	US-09-252-991A-5793
26	44	2.4	1335	4	US-09-252-991A-5877
27	43.6	2.3	7218	1	US-08-232-463-14

28	43	2.3	807	4	US-09-818-780-16	Sequence 16, Appl
29	43	2.3	885	4	US-09-252-991A-7525	Sequence 7525, Ap
30	43	2.3	969	4	US-09-252-991A-7599	Sequence 7599, Ap
31	43	2.3	1101	4	US-09-252-991A-7952	Sequence 7952, Ap
32	42.4	2.3	897	4	US-09-023-655-1332	Sequence 1332, Ap
33	42.4	2.3	1296	4	US-09-252-991A-12123	Sequence 12123, A
34	42.4	2.3	1326	4	US-09-252-991A-12394	Sequence 12394, A
35	41.2	2.2	1071	4	US-09-489-039A-3061	Sequence 3061, Ap
36	41.2	2.2	1092	4	US-09-489-039A-3067	Sequence 3067, Ap
37	41.2	2.2	2595	4	US-09-810-268-2	Sequence 2, Appli
38	41.2	2.2	3007	4	US-09-810-268-1	Sequence 1, Appli
39	41	2.2	564	4	US-09-252-991A-12263	Sequence 12263, A
40	40.8	2.2	1190	5	PCT-US91-02626-8	Sequence 8, Appli
41	40.4	2.2	402	4	US-09-489-039A-458	Sequence 458, App
42	40.2	2.2	825	1	US-08-312-870-6	Sequence 6, Appli
43	40.2	2.2	1338	1	US-08-307-444A-8	Sequence 8, Appli
44	40.2	2.2	1338	1	US-08-587-389-8	Sequence 8, Appli
45	40.2	2.2	1368	1	US-08-307-444A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-15035
; Sequence 15035, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15035
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15035

Query Match 3.5%; Score 65.6; DB 4; Length 504;
Best Local Similarity 49.3%; Pred. No. 3.7e-06;
Matches 201; Conservative 0; Mismatches 204; Indels 3; Gaps 1;

QY	688	GAGTGGTCTGCTCCCGACGCCAATCACCAGGGGAACACCTTTGGGGGCCAGATCATG	747
DB	37	GTGCTCATGACCCCGGACAGGGAACTTCTCCGGCAACGTCACGGCGGCGGCTGCTC	96
QY	748	GCCTGGATGAGAAATGGCCACCATTTGACGACGCGGCTCTGCCGTCGCCACCTTACG	807
DB	97	AAGTCTCTGACGAAGTGGCTTTGCTGCGCAAGCGCTACGCCGACGCTAGTGGTG	156
QY	808	CTGAAGGCCATTGAAATGTTCCACTTCCGAGGCGCGTCCAGGTCGCGACCGTCTGGTG	867
DB	157	ACCTGTCGGTGGATCAGGTGATCTTCCGCGAGCCGATCATGTCGGCGAACTGGTGACC	216
QY	868	CTCAAGGCCATCTGACATGCTTCAACATAGATGAGTGGCGGTGGTGAG	927
DB	217	TTCTCGCTCGGTGAACTACACCGGCGGCACCTTCAGATGAGGTGGGTAAGGTGATG	276
QY	928	GCCTATCGCGAGGCGGTGAGACCCCGCGGCCACATCAACAGTGCCTTTATGACCTTT	987
DB	277	ACCGAGACATCCAC---GAGCGCACGTCGCGCATACCAACAGTTGCTTCTCACCATG	333
QY	988	TGCTGCTGGACGACGATGACACCGCCAGTTCGTCCTCGATTCGGCCCGGCGGCGG	1047
DB	334	GTGGCGATGACGACGACCGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	393

QY 1048 GATGGTACGGCGGTACCGAGAGCGAGTGCACGAAAGAGATCCGC 1095
Db 394 ATCGAGAAACGCGCTACGCCAGGCGCTTGGCCGCGGAGCAGCGC 441

RESULT 2

US-09-252-991A-1680
; Sequence 1680, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 1680

; LENGTH: 939

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1680

Query Match 3.3%; Score 60.4; DB 4; Length 939;
Best Local Similarity 49.0%; Pred. No. 6.9e-05;
Matches 190; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

QY 689 AGTGGTCTGCTCCCAACGCAATCACCAGGGCAACACCTTTGGGGCCAGATCATGG 748
Db 308 AGGGGTATTCCCGCCCAACCAATCACCACACCTGTTCCGGGGACTGCGCTGG 367
QY 749 CTGATGAGAAATGTGGCCACCATTCAGCCAGCCGCTTCGCGTCCACCTACGC 808
Db 368 CTTGATGAGAGAGGTCTCTGTTTCATCACCGGACCCGTTTCTGCGCCCTGCGTGTGA 427
QY 809 TGAAGGCCATTGAAATGTTCCACTTCCGAGCCGCTCCAGTCCGGCCACCGTCTGGTGC 868
Db 428 CCGTTTCACCGATCGCATGCACTTCAACACCCGATCCCGGGGGGCTCGATCGTGA 487
QY 869 TCAGAGCCATC--GTGAACAATGCTTCAACATAGCATGAGGTGGCGTGTGCGTGA 926
Db 488 ACTGTCGAGCGGTGGTCAACGTCCGCAACACCATCTCAAGTGGAGTTCAGGTGA 547
QY 927 GGCCTATCCGAGGAGGTGAGACCCACCGGCGCCACATCAACAGTCCCTTATGACCTT 986
Db 548 CTGGAAGAGCATGTACGCCGACGGTTCGCGAAGGCCATACACGGCCTGTTTCAGCTTCGT 607
QY 987 TGTGTCCTGACGAGATGACCCAGCCCGCTGCTGCTGCTGATTCGGCCCGCCAGCCGG 1046
Db 608 TGCCATCGACGAGAAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667
QY 1047 CGATGGTACGCGCGGTACCGAGAGGCC 1074
Db 668 GGCTTGACAGAGGCGAAGAAAGGACACC 695

RESULT 3

US-09-252-991A-2523
; Sequence 2523, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2523
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2523

Query Match 3.1%; Score 58.4; DB 4; Length 609;
Best Local Similarity 47.1%; Pred. No. 0.00018;
Matches 213; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 656 TGCCGGCTGAGAAGACCCGTGTGGAGAGTGTGGAGCTGCTGCTCCCAACGCAATC 715
Db 128 TGCAACCCGGAACCAACCACTGTGATGACCGTGTGATGACTCGGACATGGCCAAAT 187
QY 716 ACCAGGGCAACACCTTTGGGGCCAGATCATGGCCCTGGATGGAGAATGTGGCCACCAT 775
Db 188 TCTCCGCAATGTCCATGGCGGCAACCTCTCTGAAGTATCTCGACGAAGTCTGCTATGCA 247
QY 776 CAGCCAGCGGCTGTGCGTGCCTCCACCTACCTGCTGAGGCCATTTGAAATGTCCACTTCC 835
Db 248 GCGCCAGTCTGATGCGGCTCGCTACGTGTGACCCCTGTGCTGATCAGGTGATCTTCC 307
QY 836 GAGGCCCTCCAGGTCCGGCAGCGTCTGGTGTCTCAAGCCCATCTGAACCAATGCTTCA 895
Db 308 GCGAGCCGCTGATGTGCGGCACTGTGTGACCTTCTGCTTCCGTCTCACTACACGGGC 367
QY 896 AACATAGCATGAGGTGGCGTGTGCGTGGAGGCCATTCGCCAGAGGCTGAGACCCACC 955
Db 368 GGACTTCCATGGAATCGCGTCAAGGTTCATCCGAGAACATCCGCGAGCAGTCCGT-- 425
QY 956 GGCGCCACATCAACAGTCCCTTATGACCTTGTGTGCTGCTGAGCCACATGACGAGCCCC 1015
Db 426 -GCGTCATACCAACAGCTGTTTCTTCACTGCTGCGCTGATGACGAGCGCAAGCGCG 484
QY 1016 AGTTGCTCCCTGATTCGGCCCGCCAGCGCGCGATGTTGAGCGCGGTACCGAGAGGCA 1075
Db 485 TGCCCGTCCGCGCTGAGCTGGAACCGCGGAGCAGAGCGCGCTTCGCCACAGGCC 544
QY 1076 GTCCGAGAAAGATCCGCTCGACAGGAAG 1107
Db 545 AACAAAGCGCGCAGCTGCGCGCGAGCTGGAG 576

RESULT 4

US-09-252-991A-2393/c
; Sequence 2393, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 2393

; LENGTH: 1518

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-2393

Query Match 3.1%; Score 58.4; DB 4; Length 1518;
Best Local Similarity 47.1%; Pred. No. 0.00023;
Matches 213; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 656 TGCCGGCTGAGAAGACCCGTGTGGAGAGTGTGGAGTGTGCTGCTCCCAACGCAATC 715

Db 650 TGCACCCGGAACACCACTGCTGATGACCGTCTGATGATCCGGACATGGCAATT 591
QY 716 ACAGGGCAACACCTTTGGGGGCAAGATCATGACCTGGATGGAGAATGTGGCCACCATTTG 775
Db 590 TCTCCGGCAATGTCATGGCGGACCCCTCTGAAGTATCTCGACGAAAGTGGCTATGCAAT 531
QY 776 CAGCAGCGGGCTCTGCGGTGCCACCTACGCTGAAGGCATTTGAATGTTCCATTCC 835
Db 530 GCGCCAGTCGCTATGCGGTGCTGCTGATGACCGTCTGCTGCTGCTGCTGCTGCTGCTGCT 471
QY 836 GAGGCGCGTCCAGGTGCGGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 895
Db 470 GCGAGCGGTGCTGCTGCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
QY 896 AACATGATGAGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955
Db 410 GGACTTCCATGGAATCGGCGTCAAGTCTATCACCAGAACATCCGCGAGCAGTCCGT-- 353
QY 956 GCGGCCACATCAACAGTCCCTTTATGACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
Db 352 -GCGTCATACCAACAGTGTCTTTCACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294
QY 1016 AGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1075
Db 293 TGCGCGTCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234
QY 1076 GTCCCAAGAAAGATCCCGCTCGACAGGAAG 1107
Db 233 AACAAAGCGCCAGCTGCGCGGAGCTGGAG 202

RESULT 5

US-09-252-991A-1630
; Sequence 1630, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1630
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1630

Query Match 3.0%; Score 56.6; DB 4; Length 480;
Best Local Similarity 49.6%; Pred. No. 0.00043;
Matches 173; Conservative 0; Mismatches 174; Indels 2; Gaps 1;
QY 689 AGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 748
Db 114 AGCGGTATTCCCGCGCCACCAACCAATCACCACACACCTGCTGCTGCTGCTGCTGCTGCTGCT 173
QY 749 CTGGATGGAGATGTGGCCACCATTTGAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808
Db 174 CTGGATGGAGAGTCTGCTTTCATCACCAGCAGCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 233
QY 809 TGAAGGCATTTGAATGCTTCCATCTCCAGGCGGCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 868
Db 234 CCGTTTCCACCGATCGCATTCGATTTCAACCCCGATCCCGGGGGGCTGCTGCTGCTGCTGCTGCT 293
QY 869 TCAAGGCATC--GTGAACCAATGCTTTCAACATAGCATGAGGTGGGGCTGTGCTGCTGCTGCTGCT 926
Db 294 ACTGGTGGAGCGGTGTGCTCAACGCTGCGCAACACCATCTCTCAAGGTGAGGTGAGGTGCTGCTGCT 353

QY 927 GGCTATCCAGGAGGCTGAGACCCACCGCGCCACATCAACAGTGCCTTTATGACCTT 986
Db 354 CTGGAAGACATGATGCGCGACGCTCGGAGAAAGGCCATACACGGGCTGTTCAGCTTCGT 413
QY 987 TGTGCTCTGGAGCGGATGACCGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035
Db 414 TGCCATCGACGAGAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 462

RESULT 6

US-09-149-476-179
; Sequence 179, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596

Db 654 TGATCCACCTGGTGGGCGCTTTCAGACTGCACCTGCACGGCTTTGTGCACGGAGGTGGA 713
QY 224 TGCTCAAGTGATTGACACACGCGTTCCTGTCGCGGAGAGCGCGTGGCTGGCCCT 283
Db 714 CMTAGAGCTCATGGATGAGTCCCGGGATCGTGGCTGCACGCCACTGCAAGACCAACA 773
QY 284 GTGTCAAGCTTCATGGATGACATCTATTTTGAGCACACCATTAGTGTGGACAAGTGG 343
Db 774 TCGTCAAGCTTCGTTGGAGCGCCATTATTTTCATGACAGATCAGAAAAGGTGGCTCA 833
QY 344 TGAATATCAAGGCAAGGTGAACGGGCGCTTCACTCCAGATGAGGTGGGATCCAGG 403
Db 834 TCACCTCTCGGACGCGATGACCTTCACGAGCAATAAGTCCATGGAGATCGAGGTGG 893
QY 404 TGGCCTCGGAGGACCTGT-----GCTCTCAGAGAGTGAATGTGTGCAAGGCTTGG 457
Db 894 TGAGCGCGGACCTGTGTGTGAGACGCTCTCAGAAAGCGCTACCGGGCGCGAGTGCCTTCT 953
QY 458 CCACCTTCTGTGGCGCGCGAGAGATCACCAAGGTGAAGCTGAAGCGAGATCACCGCGCGGA 517
Db 954 TCACCTACGTGTGCTGAGCGAGGAGGAGTGTGCTGTGCCCCAGCTGGTGGCCCG 1013
QY 518 CAGAAGAGGAGATGAGGACACAGTGTGGCGGTGAGCGCGCGCGCATGC 568
Db 1014 AGACCGAGGACGAGAGAAGCGCTTTGAGGAAGGCAAGGGCGGTACCTGC 1064

RESULT 7

US-09-149-476-306
; Sequence 306, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: F2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888

US-09-328-111-475

Query Match 2.7%; Score 50.6; DB 3; Length 647;
Best Local Similarity 86.2%; Pred. No. 0.011; Indels 0; Gaps 0;
Matches 56; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1793 TCCCGGTGGCCAAATCAACGCGGATTTGATACCTTCAAGACACCTGAAACCTTATCA 1852
DB 289 TTCTCCAGGCCATATCATATAGCTATTGATACCATCAAGACACCTGAAACCTTATCG 348
QY 1853 TGAGC 1857
DB 349 TGAGC 353

RESULT 12

US-09-252-991A-5935
; Sequence 5935, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5935
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5935

Query Match 2.6%; Score 49.2; DB 4; Length 975;
Best Local Similarity 45.5%; Pred. No. 0.027;
Matches 174; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 1176 CTTGAGCTACATAACGCTCTCTCTTGAAGATGCTTGTGGCCAAAGGCAACTGGGTGCT 1235
DB 67 CCAGGCCACGACGCGCTGCTCAAGGCTTCTCTGAGGAAATCGGTACTCTGCTGCC 126
QY 1236 GTCCTCGGAGATCAGTCAAGTCCGCTGTACACTCTGGAGGATGCAAGTTCCTCTCTT 1295
DB 127 CGAGCGGAAGACTTCCAGGCGCGCACCCAGAACTGCGACGAGATGCGCCGATGCG 186
QY 1296 CCACATGAGATGTTGGTGCATGTGATGCGAGCCGCTTCTCTGCTGCTCTCGACCT 1355
DB 187 CGGCCGCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 246
QY 1356 GCGTCAGAGCCAGAGTGGGACACGACTACCGAGAGTGGAGTGTGACAGGATGACAGGTAGA 1415
DB 247 CGCTGGGGCTCGCTGTACAGGCACTTACGGCCACGACGCTGATCAGCGAAGAGGCGG 306
QY 1416 CGAGCAGCAGCCATCTACAGTCCAGGCTCCGCTCGAGGCTCACACAAAGCCCA 1475
DB 307 CGCCGAGAAAGCAGGGCTACAAAGGTTCCGGCGGCAAGGTCATCGCTTCGCTCG 366
QY 1476 GGACTTCGTGATCTTGGCTCGAGGCGGAAGCTTTGTGACAAATGGGACCCCTATGTCTAT 1535
DB 367 CGCTTCCTCGAGGCGCGCGCTGGAGTCCGCTCCCATGTGCGAGGCCACTTCCTTA 426
QY 1536 CGCGCTGAGTGGTCCGCTCAGCTG 1557
DB 427 CAGCGTGAAGACGGCGCGCTG 448

RESULT 13

US-09-252-991A-5953
; Sequence 5953, Application US/09252991A

Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5953
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5953

Query Match 2.6%; Score 49.2; DB 4; Length 2196;
Best Local Similarity 45.5%; Pred. No. 0.033;
Matches 174; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 1176 CTTGAGCTACATAACGCTCTCTCTTGAAGATGCTTGTGGCCAAAGGCAACTGGGTGCT 1235
DB 237 CCAGGCCACGACGCGCTGCTCAAGGCTTCTCTGAGGAAATCGGTACTCTGCTGCC 296
QY 1236 GTCCTCGGAGATCAGTCAAGTCCGCTGTACACTCTGGAGGATGCAAGTTCCTCTCTT 1295
DB 297 CGAGCGGAAGACTTCCAGGCGCGCACCCAGAACTGCGACGAGATGCGCCGATGCG 356
QY 1296 CCACATGAGATGTTGGTGCATGTGATGCGAGCCGCTTCTCTGCTGCTCTCGGACCT 1355
DB 357 CGGCCGCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 416
QY 1356 GCGTCAGAGCCAGATGGGACACGACTACCGGAGCTGGAGTGTGACAGGATGACAGGTAGA 1415
DB 417 CGCTGGGGCTCGCTGTACGACGCACTCTACGGCCACGACGCTGATCAGGAAGAGCGG 476
QY 1416 CGAGCAGCAGCCATCTACAGCTCACCAGGCTCCGCTCGAGGCTCACAAAGCCCA 1475
DB 477 CGCCGAGAAAGCAGGGCTACAAAGGTTCCGGCGGCAAGGTCATCGCTTCGCTCG 536
QY 1476 GGACTTCGTGATCTTGGCTCGAGGCGGAAGCTTTGTGACAAATGGGACCCCTATGTCTAT 1535
DB 537 CGCTTCCTCGAGGCGCGCGCTGGAGTCCGCTCCCATGTGCGAGCCACTTCCTTA 596
QY 1536 CGCGCTGAGTGGTCCGCTCAGCTG 1557
DB 597 CAGCGTGAAGAACGGCGCGCTG 618

RESULT 14

US-09-252-991A-5934/G
; Sequence 5934, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5934
; LENGTH: 3036
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5934

Query Match 2.6%; Score 49.2; DB 4; Length 3036;
Best Local Similarity 45.5%; Pred. No. 0.036;
Matches 174; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 1176 CCTGAGCTACAAATACGCTCTCTCTTGAAGATGCTTGTGCGCAAGGACAACTGGTGCT 1235
DB 2068 CCAGGCCACAGCGCGCTGCTTACAGGCCCTTCTCGAGGAAATCGGCTACCTGCTGCC 2009
QY 1236 GTCCTCGGAGATCAGTCAGGTCGCGCTTACACTCTGGAGATGACAAGTTCTCTCTCT 1295
DB 2008 CGAGCGGAACACTTCCAGGCGGACCCAGAAAGTCGACGACGAGATCGCCCGCATGGC 1949
QY 1296 CCACATGAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1355
DB 1948 CGGCCCGCAGCTGT 1889
QY 1356 GCGTCAGAGCCAGAGTGGGACAGCACTACCGGAGCGTGGAGTAGTGGACAGGTAGA 1415
DB 1888 CCGCTGGGGTCTCGTGTACGACGCACTCTACGCGACCGACGTGATCAGCGAAGAGCGG 1829
QY 1416 CGAGGACGACGCCATCTACACGCTCACCAGCCCTCGGAGGTGACAAAGGCCCA 1475
DB 1828 CGCGAAGAGGCAAGGCTTCAACAGGTTCCGCGGCAAGGTCAATCGCTTCGCTCG 1769
QY 1476 GGAATTCGTGATCTCGGCTCGAGCGGAAAGCCTTGTGACAAATGGGACCCCTATGTCA 1535
DB 1768 CGCTTCTCGACGAGCGCGCGCTGGAGTCCGGTCCCATGTCGACGCCACTTCCTA 1709
QY 1536 CGCGCTGAGGTGCTACGGTG 1557
DB 1708 CAGCGTGAAGACGGCGCGCTG 1687

RESULT 15
US-09-252-991A-14378
; Sequence 14378, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14378
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14378

Query Match 2.6%; Score 48; DB 4; Length 270;
Best Local Similarity 50.4%; Pred. No. 0.036;
Matches 117; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 374 TCAATCCAGCATGAGGTGGGCATCCAGTGGCTCGGAGGACCTGTGCTGTGAGAGC 433
DB 16 TCGACCGCGCGGGTCTTCGCCACCTGTGTGGCACCTTGGAACTGCTTCGACTCGG 75
QY 434 AGTGGATGTGCAAGGCTTGGCCACCTTGTGGCCCGCGAGAGATCAACAAAGTGA 493
DB 76 AGCGGAGCGGGTCAAGGGCTTGTGTCATCAATCGCTTCGCGCGGACATCGCCCTGCTGC 135
QY 494 AGCTGAAGCAGATCAGCGCGGACAGAGAGAGATGAGACAGTGTGGCGGCTG 553
DB 136 AACCCGGCTTCGACTGGCTGGAGGACGACCGGCAAGCGGTGCTCGCGGTGCTGCCCT 195
QY 554 AGCGCGGCGCATGCGCTTGTGTATGACAGACCATCAAGGACCTTCCTGGC 605

DB 196 ATGTACGCACTGCACCTGGAGCGGAGGACGCCATCGACACGCGCAGGC 247

Search completed: August 21, 2004, 01:01:42
Job time : 154 secs

Blank Sheet